LHCb: Vertex detector and read out chip, Computing challenge and B-> J/ psi phi analysis (zie titelpagina)

Klous, S.

2005

document version
Publisher's PDF, also known as Version of record

VRIJE UNIVERSITEIT

*LHC*b:
Vertex detector and read out chip,
Computing challenge
and $B_s \rightarrow J/\psi\phi$ analysis

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op dinsdag 31 mei 2005 om 10.45 uur
in het auditorium van de universiteit,
De Boelelaan 1105

door
Sander Klous
geboren te Zaandam

promotor:	prof.dr.ing. J.F.J. van den Brand
copromotoren:	dr. E. Jans
	dr. H.G. Raven

*LHC*b:
Vertex detector and read out chip,
Computing challenge
and $B_s \rightarrow J/\psi\phi$ analysis

Cover design: M.J. Kraan.

ISBN 90-738-3843-6.
Produced by F&N Boekservice, Amsterdam.

The work described in this thesis is part of the research programme of the "Stichting voor Fundamenteel Onderzoek der Materie" (FOM), which is financially supported by the "Nederlandse Organisatie voor Wetenschappelijk Onderzoek" (NWO).

Contents

Chapter 1

Introduction

One of the remaining mysteries in cosmology is the absence of antimatter in our universe. The explanation of this deficit may be found in the behaviour of elementary particles. Starting with the generally accepted assumption that equal amounts of matter and antimatter were once present in our universe, one is left with the conclusion that differences in the behavior of matter and antimatter must have led to this baryon asymmetry. The *LHC*b experiment, as discussed in this thesis, will investigate these differences. It focuses specifically on the violation of the combined Charge and Parity symmetries in hadrons containing b and $\overline{b}$ quarks. The Standard Model accommodates this violation within the weak interaction by means of a difference between the weak and the mass eigenstates of the three quark generations.

*LHC*b will be installed in the Large Hadron Collider (LHC) and is scheduled to commence operation in 2007. In the proton$-$proton collisions of this accelerator an enormous amount of b-hadrons will be produced due to the high luminosity and the high energy of the interactions. However, the collision rate, the radiation levels and the high energy of the particles involved provide an extremely hostile operating environment for the detectors of *LHC*b. This environment, in combination with the required accuracy to put significant constraints on (Standard Model) observables, results in a stringent set of specifications for the sub-detectors. These specifications include fast electronics, high precision mechanical structures, ultra high vacuum systems and high speed data processing. Hence, state of the art technologies in mechanical and electrical engineering, as well as in computer science are required to achieve the scientific goals of *LHC*b.

All these technologies are implemented in the vertex detector, which directly surrounds the interaction point and plays a crucial role in the level-1 trigger of *LHC*b. The electrical components need to comply with the requirements as specified for this trigger, *e.g.* pulse shape, latency and output rate. Furthermore, the proximity of the detectors to the interaction point results in a high radiation dose for sensors and readout electronics, which introduces complications in their design and operation. Here, a prototype of the chips that read out the signals from the hundreds of thousands of silicon strips in the vertex detector and other sub-detectors (pile-up detector, trigger tracker and inner tracker) is discussed. These chips should be able to process the detector signals fast enough to separate pulses associated with different bunch crossings. Their compliance with the specifications needs to be demonstrated in a realistic setup, since their performance could seriously be degraded by the operating environment.

Mechanically, the challenge in the design of the vertex detector is to find a compromise between two contradicting requirements. On the one hand, from the point of view of reliability and operational issues, one would like to separate the detector system from the circulating beams. On the other hand, with respect to physics, any material between the sensors and the interactions disturbs the measurements.

From a computer science point of view, the handling of the enormous amount of data that the LHC experiments will produce is considered one of their most severe operational challenges. In response to this challenge a huge amount of effort and resources is being invested in developments in the field of information technology. Besides the required technological progress, these investments also aim at organizational and political benefits, leading to the construction of large-scale distributed computing infrastructures. These infrastructures, known as grids, combine many resources of the contributing universities and institutes. Adoption of grid computing has so far been limited to computing experts, running programs specifically tailored to the grid environment. Here, the development of a comprehensive user interface is addressed, which facilitates deployment of (grid-unaware) applications on a distributed infrastructure by a broader user base.

All the equipment together eventually needs to provide the physics performance required by *LHC*b. This physics performance is studied with simulations during the design and construction of an experiment. In the first stage everything from physics events to expected detector responses is simulated based on the experiment specifications. As developments continue, detailed information from all sorts of (beam) tests are incorporated, to ensure that the results are as realistic as possible. Results of these simulations led to an extensive redesign of LHCb between the Technical Proposal [1] and the Re-optimization TDR [2] to drastically reduce the amount of material in the detector.

With the simulated detector responses Monte Carlo data is analyzed as if the experiment is already up and running. All sorts of observables (like reconstructed masses and transverse momenta), as made available by the reconstruction and analysis software, are combined to optimize the selection of specific decays to obtain the best possible signal efficiency and purity. Note that the quality of the available observables is directly related to the mechanical and electrical performance of the various sub-detectors. With the selected data sample the expected sensitivity of the experiment for a variety of physics parameters is determined, which allows for a direct comparison to the required physics performance.

This thesis can roughly be divided in four parts: theory, detector engineering and testing, grid computing, simulation and analysis. Each of these parts is briefly described below.

In chapter 2 the theoretical context of the *LHC*b physics program is discussed, including some details about the behavior of b-mesons. The $B_s \rightarrow J/\psi\phi$ decay is specifically addressed, demonstrating the possible implications of this measurement for the Standard Model description of CP violation.

In chapter 3 some of the engineering aspects of the *LHC*b experiment are presented with emphasis on the challenges for the trigger, the track reconstruction and particle identification. In the last section of this chapter a detailed description of the vertex detector is given: layout, vacuum system, operational procedures, silicon sensors and cooling of this sub-detector are described. The section is concluded with a discussion of the readout chip of the silicon sensors, called 'the Beetle'. In chapter 4 a test at the X7 beam test facility at CERN is described with a prototype hybrid for the vertex detector, equipped with 16 Beetle chips.

The results are translated into a threshold scan, providing the opportunity to tune the settings for maximal signal efficiency and purity. Furthermore, aspects specific for the actual *LHC*b running conditions are mimicked with dedicated measurements.

In chapter 5 the *LHC*b Monte Carlo production system and its deployment on a large-scale distributed computing system is presented. Next, several software components are combined, allowing users to almost transparently use grid resources as if they were locally available while providing transparent access to files, even when private networks intervene. These ideas are implemented for BaBar Monte Carlo production software using the European Data Grid. They can however be realized in a variety of environments.

Chapter 6 contains an overview of the simulation, reconstruction and analysis software involved in the *LHC*b performance studies, followed by a discussion of the optimization of the selection of events that contain $B_s \rightarrow J/\psi\phi$ decays. For this purpose, a new algorithm is developed that takes any set of observables and identifies the criteria to obtain the global optimum of signal efficiency and purity. The last part of chapter 6 describes the simulation of physics results that can be extracted from the $B_s \rightarrow J/\psi\phi$ analysis by generating the relevant observables from their appropriate probability density functions with a so-called Toy Monte Carlo study. The chapter is concluded with an update of the results of this study for a nominal year of data taking of the *LHC*b experiment.

1 Introduction

Chapter 2

Theory

The universe appears to be populated exclusively by matter rather than by antimatter [3]. The production of this net baryon asymmetry requires baryon number violating interactions, departure from thermal equilibrium and violation of the symmetries C (charge conjugation) and CP (charge conjugation combined with parity transformation), as described by Sakharov in 1967 [4]. Here, a possibility of CP violation is addressed in the context of the Standard Model.

2.1 Introduction to the Standard Model

The Standard Model is a theoretical framework, which describes interactions between elementary particles. It includes the strong force and the electroweak force, but it does not include the effects of gravitation. All phenomena as observed so far can either be described by this model or by general relativity [5] with high precision, although the ultimate goal of many physicists is a unified theory that includes all forces of nature[1].

The elementary particles of the Standard Model are pointlike, without substructure and can be classified into two categories. The first category contains the fermions obeying Fermi statistics. Translated into fundamental properties, this makes that their spins are half integer. The fermions can be subdivided in two groups, leptons and quarks. In contrast to leptons, quarks interact through the strong force and therefore have an additional quantum number, color, to describe their behavior. Color can be of three types (red, green or blue). However, it must be a 'confined' property, since colored objects are only observed in composite particles. These composite particles are either mesons (in case of a color − anti-color pair, *i.e.* a quark−antiquark combination) or hadrons (in case of a red − green − blue combination, *i.e.* containing three quarks).

The second category of particles of the Standard Model contains bosons, which obey Bose-Einstein statistics and have integer spin. The mediators of the strong and electroweak forces are known to be bosons. The strong force can be described by the exchange of eight kinds of massless gluons, connecting the different colors. The electromagnetic part of the

[1]Today, the prime candidate for such a unified theory is string theory. So far, no direct observations exist that indicate nature should be described by such a theory. The firm believe in its existence is primarily based on extrapolation of known phenomena and on esthetics.

electroweak force is carried by the massless photon and the weak part by its heavy counterparts the $W^{\pm}$ and Z bosons[2]. An overview of all leptons and quarks in the standard model, together with the mediators of the electroweak and strong force is given in Fig. 2.1.

Fermions				Gauge bosons		
Charge	Name					
Quarks $+\frac{2}{3}$	u *up*	c *charm*	t *top*	Name	Charge	Force
Quarks $-\frac{1}{3}$	d *down*	s *strange*	b *bottom*	g *gluon*	0	Strong
Leptons 0	ν_e *electron neutrino*	ν_μ *muon neutrino*	ν_τ *tau neutrino*	γ *photon*	0	EM
Leptons -1	e *electron*	μ *muon*	τ *tau*	Z *boson*	0	Weak
				W *boson*	± 1	Weak
Generation	I	II	III			

Figure 2.1: Overview of the lepton and quark sector of the Standard Model. The bosons mediating the electroweak and strong force are shown as well.

The ideas described above are formalized mathematically by means of quantum field theory, based on gauge symmetries. Furthermore, the gauge invariant theories that constitute the Standard Model are renormalizable, which means that they allow for quantitative predictions of experimental results [6]. The details of the described interactions find their origin in the properties of symmetry groups. The simplest example is quantum electrodynamics (QED), which contains a manifest gauge symmetry. The Lagrangian is invariant under local transformations of the phase of the wave function. The corresponding local symmetry group for these transformations is $U(1)$. The Maxwell equations as well as conservation of electric charge can be deduced directly from these symmetry relations.

The strong force with its three colors requires an extension of the phase transformation symmetry to color space. Hence, a vectorized phase transformation describes the underlying $SU(3)_{Color}$ symmetry, which leads to a non-Abelian theory called Quantum Chromo Dynamics (QCD). The group properties of this theory result in self-interaction of gluons, color confinement and asymptotic freedom of the quarks [7].

Quantum electrodynamics and the strong force are both based on unbroken gauge symmetries, which implies that photons and gluons must be massless (a mass terms would violate gauge invariance of the Lagrangian). The integration of weak interaction with QED on the other hand, is based on a spontaneously broken symmetry as proposed by Weinberg [8] and Salam [9]. The model combines a Lagrangian, invariant under $SU(2)_{Left}$ (transforming only the left handed electron and neutrino) and $U(1)$ (hyperspace) transformations, with a scalar field. The specific potential of the scalar field introduces a new (degenerate) ground

[2]The particle mediating gravitation is thought to be the graviton, a boson with a spin of 2.

state, where the symmetry of the described system appears to be reduced and the originally massless bosons become massive, a phenomenon referred to as the Brout−Englert−Higgs mechanism [10, 11]. However, the resulting field equations describing the vacuum states are coupled, which means that the mass matrix in the Lagrangian first needs to be diagonalized to allow for an interpretation as physical fields. After diagonalizing this matrix, one combination of fields remains massless and it can be shown that the associated particle fulfils the Maxwell equations. As a result, this particle is identified as the photon. The other field combinations give rise to the massive $W^{\pm}$ and Z^0 bosons and provide a description of the weak interaction. 't Hooft and Veltman proved that theories with spontaneously broken gauge symmetries remain renormalizable [12], which enables the accurate calculation of electroweak processes.

The resulting Lagrangian of the electroweak interactions is shown in equation (2.1).

$$\begin{aligned}\mathcal{L}_{EW} = \; & +g\cdot\sin\theta_W\overline{e}\gamma^\mu eA_\mu + i\left(\overline{\nu}\gamma^\mu\partial_\mu\nu + \overline{e}\gamma^\mu\partial_\mu e\right) - \overline{e}e\cdot m_e \\ & + \tfrac{g}{\sqrt{2}}\left(\overline{e_L}\gamma^\mu\nu W_\mu^- + \overline{\nu}\gamma^\mu e_L W_\mu^+\right) + \tfrac{g^2}{4}\left(W_\mu^+W^{-\mu} + \tfrac{Z_\mu Z^\mu}{2\cdot\cos^2\theta_W}\right)m_{W^\pm}^2 \\ & + \tfrac{g}{\cos\theta_W}\left(\tfrac{1}{2}\overline{\nu}\gamma^\mu\nu - \tfrac{1}{2}\cos 2\theta_W\cdot\overline{e_L}\gamma^\mu e_L + \sin^2\theta_W\overline{e_R}\gamma^\mu e_R\right)Z_\mu + \mathcal{L}_{kin.}.\end{aligned} \tag{2.1}$$

This equation contains the electron (e), a massless neutrino (ν), a massless photon field (A_μ), the massive boson fields ($W^{\pm}$ and Z) and the lepton − boson interactions. The subscript L (R) denotes left-handed (right-handed), θ_W is the weak mixing angle, m is mass and g is the electroweak coupling constant. The kinetic terms of the gauge fields are omitted, since they are not of interest for the discussion which follows. The quarks are introduced in the next section, where the extension of this equation forms the basis of the description of CP violation in the electroweak sector of the Standard Model.

2.2 CP violation in the weak interaction

The mass term of the electron in equation 2.1 would break the gauge invariance of the Lagrangian. However, it can consistently be introduced via a coupling to the Higgs field. This so-called Yukawa coupling is gauge invariant, although this is not made explicit in the extended Lagrangian of equation (2.2).

$$\begin{aligned}\mathcal{L}_{EW} = \; & +g\cdot\sin\theta_W\left(\overline{l}\gamma^\mu l + \tfrac{2}{3}\overline{u}\gamma^\mu u + \tfrac{1}{3}\overline{d}\gamma^\mu d\right)A_\mu \\ & + i\left(\overline{\nu}\gamma^\mu\partial_\mu\nu + \overline{l}\gamma^\mu\partial_\mu l + \overline{u}\gamma^\mu\partial_\mu u + \overline{d}\gamma^\mu\partial_\mu d\right) \\ & - \left(\overline{l_L}M_l l_R + \overline{u_L}Mu_R + \overline{d_L}M_d d_R + h.c.\right) \\ & + \tfrac{g}{\sqrt{2}}\left(\overline{l_L}\gamma^\mu\nu W_\mu^- + \overline{\nu}\gamma^\mu l_L W_\mu^+ + V^{ij}\overline{u_{Li}}\gamma^\mu d_{Lj}W_\mu^- + V_{ij}^*\overline{d_L^j}\gamma^\mu u_L^i W_\mu^+\right) \\ & + \tfrac{g}{\cos\theta_W}\left(\tfrac{1}{2}\overline{\nu}\gamma^\mu\nu - \tfrac{1}{2}\cos 2\theta_W\cdot\overline{l_L}\gamma^\mu l_L + \sin^2\theta_W\overline{l_R}\gamma^\mu l_R\right)Z_\mu \\ & + \tfrac{g}{\cos\theta_W}\left(\left(\tfrac{1}{2} - \tfrac{2}{3}\sin^2\theta_W\right)\overline{u_L}\gamma^\mu u_L + \left(-\tfrac{1}{2} + \tfrac{1}{3}\sin^2\theta_W\right)\overline{d_L}\gamma^\mu d_L\right)Z_\mu \\ & - \tfrac{g}{\cos\theta_W}\left(\tfrac{2}{3}\sin^2\theta_W\overline{u_R}\gamma^\mu u_R - \tfrac{1}{3}\sin^2\theta_W\overline{d_R}\gamma^\mu d_R\right)Z_\mu \\ & + \tfrac{g^2}{4}\left(2W_\mu^+W^{-\mu} + \tfrac{Z_\mu Z^\mu}{2\cdot\cos^2\theta_W}\right)m_{W^\pm}^2.\end{aligned} \tag{2.2}$$

The masses for the other leptons and for the quarks are introduced in a similar manner together with the couplings of these particles to the photon and the $W^{\pm}$ and Z^0 bosons. The following notation is used: $l = \begin{pmatrix} e \\ \mu \\ \tau \end{pmatrix}$, $\nu = \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix}$, $u = \begin{pmatrix} u \\ c \\ t \end{pmatrix}$ and $d = \begin{pmatrix} d \\ s \\ b \end{pmatrix}$.

There is no apparent reason to assume that the mass matrices are diagonal for the eigenstates of the electroweak couplings. Since any square matrix can be diagonalized with two unitary matrices, all quark mass terms are extended to $\overline{u_L} U_L M_u {U_R}^\dagger u_R$ and $\overline{d_L} U'_L M_d {U'_R}^\dagger d_R$. In other words, the quark states are redefined to match the diagonal mass matrix (*e.g.* $\overline{u_L} \rightarrow \overline{u_L} U_L$). By redefining these quark states, the electroweak couplings are affected as well. Note that due to unitarity ($U_L U_L^\dagger = U_R U_R^\dagger = 1$), flavor changing interactions in the electromagnetic and Z^0 couplings do not appear. Higher order flavour changing neutral interactions in the Standard Model are suppressed by the so-called GIM mechanism [13], which states that loop diagrams would cancel each other if quark masses would have been equal. In the charged current coupling to the $W^{\pm}$ bosons on the other hand, a unitary matrix $V^{ij} = U_L {U'_L}^\dagger$ appears, which allows to couple different flavors. Note that only left-handed quarks couple to the $W^{\pm}$ bosons, just as with leptons. The introduction of V^{ij}, also known as the CKM matrix (named after Cabibo, Kobayashi and Maskawa) [14, 15], offers the opportunity to break the CP invariance of the Lagrangian (if $V^{ij} \neq V^{ij^*}$) in the case that three or more quark families exist[3]. This symmetry breaking was already discovered experimentally in 1964 by J.H. Christenson *et al.* [16], by showing that the assumed CP-eigenstate K_{long} is not entirely CP-odd, since it was observed to decay into the CP-even final state $\pi^+\pi^-$. In the rest of this section the properties of the CKM matrix are discussed in more detail.

First of all, in case of 3 families, the 9 complex numbers of the matrix provide 18 degrees of freedom. However, there are 9 unitarity conditions and equation (2.2) shows the possibility to re-phase all relative quark states without physical consequences (absorbing 5 degrees of freedom). The standard way to parametrize the remaining 4 degrees of freedom is based on the mixing angles between the different families, as shown in equation (2.3) (with $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$). Note that in addition to the three mixing angles one complex phase $\left(e^{\pm i\delta}\right)$ arises. It is this phase that provides the opportunity to break CP invariance in the Standard Model.

$$\begin{aligned} V_{CKM} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}. \end{aligned} \tag{2.3}$$

The unitarity of the CKM matrix implies various relations between its elements. The 6 off-diagonal combinations of $V_{CKM} V_{CKM}^\dagger$ allow for an interesting visualization of these relations in the form of so-called unitarity triangles. The lengths of the sides of these triangles

[3]This led to the prediction that three undiscovered particles existed, since in 1973 only the up, down and strange quarks were known.

can be determined from decay rates, which are associated to tree-level processes involving CKM matrix elements. These processes demonstrate the existence of a clear hierarchy in the matrix coefficients. The absolute values of the diagonal elements are close to one and further from the diagonal the elements are smaller, *i.e.* transitions between different families are suppressed. Wolfenstein makes use of this hierarchy by expanding the matrix elements in terms of $\lambda = \sin\theta_{12} \approx 0.22$ and $A\lambda^2 = V_{cb} \approx 0.04$ [17]. Unitarity then prescribes the following form, where two more parameters (ρ and η) have been introduced,

$$V_{CKM} = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-\eta) & -A\lambda^2 & 1 \end{pmatrix} + O\left(\lambda^4\right). \tag{2.4}$$

The Standard Model does not require the suppression of transitions between different families (*i.e.* besides the unitarity constraints, the matrix elements are free parameters), which means this hierarchy is probably a manifestation of a more fundamental theory. The large number of measurable parameters and their relations allows for internal consistency checks of the Standard Model and for stringent tests of the predictions made by more fundamental models. The CKM matrix becomes even more interesting with recent evidence that neutrinos oscillate and have mass. These neutrino properties indicate the existence of a similar matrix in the lepton sector, which seems to exhibit a very different hierarchy.

Hence, it is interesting to measure the CKM parameters precisely. One way to analyze the results of these measurements is by investigating the shape of the unitarity triangles. Only two of them contain sides which are all of the same order in λ, these are shown in Fig. 2.2. The triangle on the left hand side combines the column of down quarks (V_{id}), with the one of bottom quarks $\left(V_{ib}^*\right)$. The triangle on the right hand side combines the row of up quarks (V_{ui}^*), with the one of top quarks (V_{ti}). Note that knowledge of relative phases between CKM elements is required to measure the angles of the triangles. When several amplitudes contribute to a decay these relative phases become observable as a result of the interference between the processes involved [18].

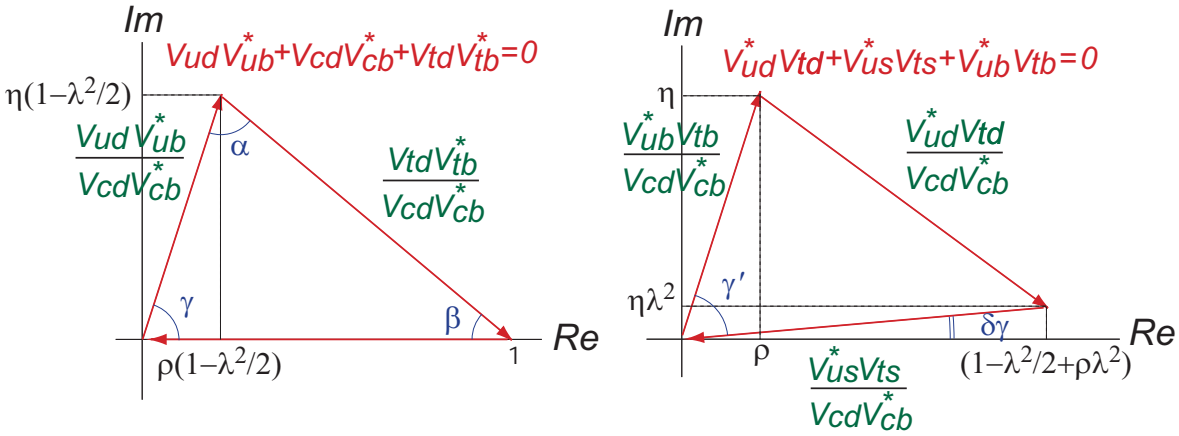


Figure 2.2: Two of the triangles visualizing off-diagonal unitarity relations of the CKM matrix. The triangle on the left combines the column of down quarks (V_{id}), with the one of bottom quarks $\left(V_{ib}^*\right)$. The triangle on the right combines the row of up quarks (V_{ui}^*), with the one of top quarks (V_{ti}).

At first, it seems that a lot of knowledge of the matrix elements is required to check the geometry of these triangles. However, the four elements in the upper left corner of the CKM matrix are tightly constrained by the Wolfenstein parametrization and can in good approximation be substituted by their expansions up to λ^2. This simplification results in the following expressions,

$$\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} = \frac{\left(1-\lambda^2\right)V_{ub}^*}{\lambda V_{cb}^*} \quad \text{and} \quad \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = \frac{V_{ub}^*}{\lambda V_{cb}^*}. \tag{2.5}$$

$$\frac{V_{ub}V_{tb}^*}{V_{cd}V_{cb}^*} = \frac{V_{ub}^*}{\lambda V_{cb}^*}, \quad \frac{V_{ud}V_{td}^*}{V_{cd}V_{cb}^*} = \frac{\left(1-\lambda^2\right)V_{td}^*}{\lambda V_{cb}^*}, \quad \frac{V_{us}V_{ts}^*}{V_{cd}V_{cb}^*} = \frac{V_{ts}^*}{V_{cb}^*}. \tag{2.6}$$

The remaining elements of interest are V_{ub}, V_{cb}, V_{td} and V_{ts}.

- The elements V_{ub} and V_{cb} involve couplings between b quarks and lighter quarks, which makes it obvious that B-mesons are required to determine these elements. An example is given by Gronau, London and Wyler [19] with the extraction[4] of arg $\left(V_{ud}V_{ub}^*/V_{cs}V_{cb}^*\right)$ from the decays $B^\pm \to D^0 K^\pm$ and $B^\pm \to \overline{D^0} K^\pm$. The interference between these two decays is introduced by considering the common final state[5] $\pi^+\pi^-$ of D^0 and $\overline{D^0}$. The two tree level Feynman diagrams, involved in the analysis are shown in Fig. 2.3. First results of these measurements are published in [20, 21].

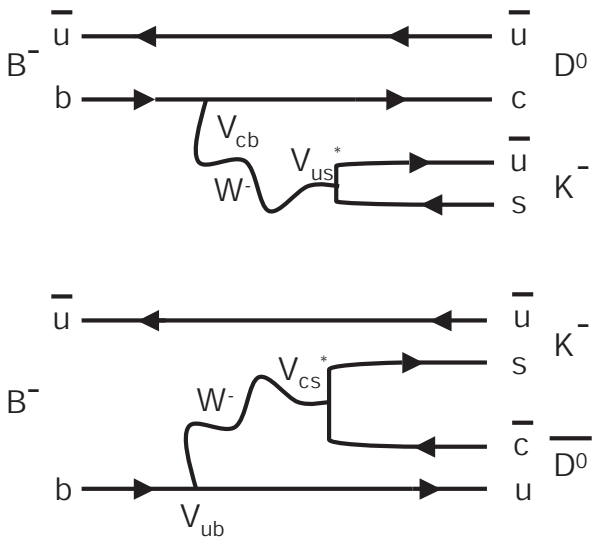


Figure 2.3: Feynman diagrams of the decays $B^- \to D^0 K^-$ and $B^- \to \overline{D^0} K^-$.

- The elements V_{td} and V_{ts} involve couplings between top quarks and down or strange quarks. Prompt decay products from the heavy top quark contain extremely high momenta, which could in principle provide the means to identify the final state products

[4] Note that $\arg V_{cs} \approx \arg V_{cd}$, so this argument provides a measure for the angle γ.

[5] Since $D^0_{CP=+} = D^0 + \overline{D^0}$ and $\pi^+\pi^-$ is a CP even final state, both D^0 and $\overline{D^0}$ can decay to this state.

and determine these couplings directly. However, almost all top quarks decay to b-quarks ($V_{tb} \approx 1$), thus the signals will be exceedingly small $\left(< O\left(\lambda^4\right)\right)$ and therefore virtually impossible to extract from the background. Here, the focus will be on indirect measurements, based on processes with top quark dominated loops. These processes are described in the next section.

2.3 Box diagrams

Some processes with top quark dominated loops are presented in Fig. 2.4. The processes shown are of special interest, since they are the lowest order Feynman diagrams capable of turning a neutral meson into its own antiparticle via the weak interaction. As a result, oscillations between particle and antiparticle states can occur. Note that the D^0 diagram does not contain the required top quark couplings as mentioned in the previous section and is therefore not discussed.

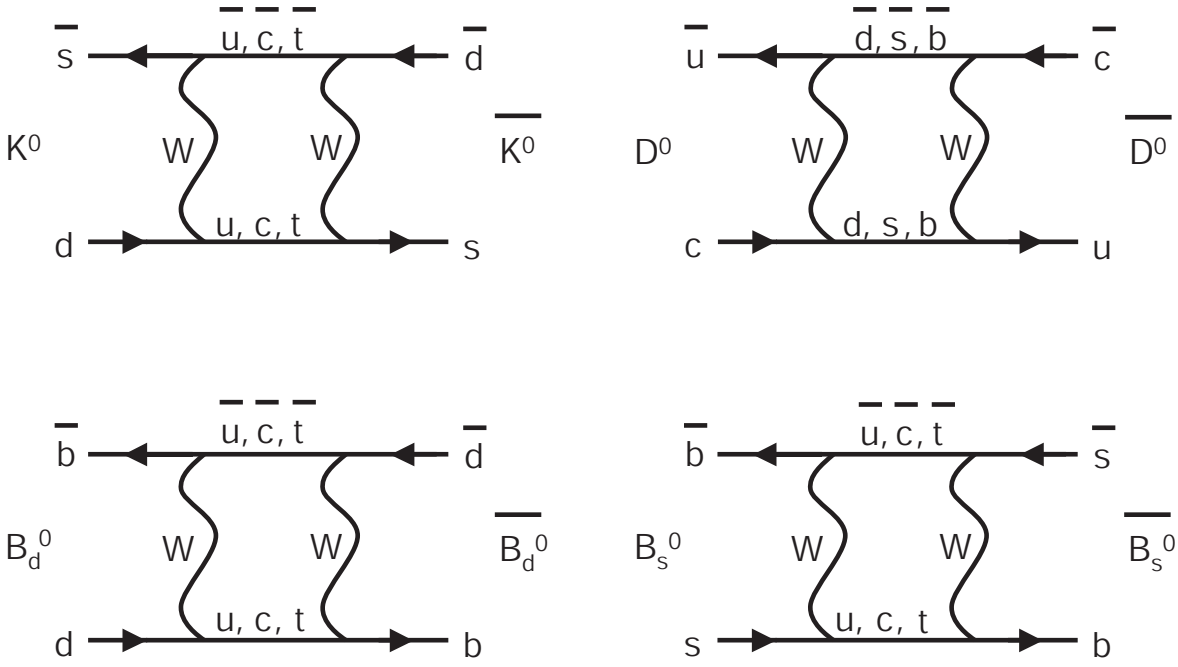


Figure 2.4: Box diagrams of K^0, D^0, B_d^0 and B_s^0.

In a quantum mechanical system, oscillations between two states can be associated with a mass difference between them. The time evolution of a two component wave function, $|\psi(t)\rangle = a(t)\left|P^0\right\rangle + b(t)\left|\overline{P^0}\right\rangle$, can be described with the following Schrödinger equation,

$$i\frac{d}{dt}\begin{pmatrix} a \\ b \end{pmatrix} = \left(\mathrm{M} - \frac{i}{2}\Gamma\right)\begin{pmatrix} a \\ b \end{pmatrix}. \tag{2.7}$$

The Hamiltonian consists of a mass matrix (M) and a decay matrix (Γ). The states $\left|P^0\right\rangle$ and $\left|\overline{P^0}\right\rangle$ develop independently when these matrices are diagonal. Otherwise, these states can mix, which gives rise to oscillations. The eigenstates of the Hamiltonian are

$p\left|P^0\right\rangle \pm q\left|\overline{P^0}\right\rangle$, *i.e.* p and q are the fractions of the original states (described by a and b), that contribute to the two eigenstates. Decay and oscillations of such a system are described by Ref. [22]

$$P\left(\left|P^0\right\rangle \rightarrow \left|\overline{P^0}\right\rangle\right) = \left|\frac{q}{p}\left(e^{-\Gamma_1 \cdot t} + e^{-\Gamma_2 \cdot t} - e^{-\frac{\Gamma_1+\Gamma_2}{2}t}\cos\left((m_2-m_1)t\right)\right)\right|^2. \tag{2.8}$$

The masses (m_1 and m_2) and decay widths (Γ_1 and Γ_2) in equation (2.8) are the properties of the two eigenstates. The mass difference between the two eigenstates is especially interesting, since it is proportional to the oscillation frequency. Hence, by calculating the diagrams of Fig. 2.4, a prediction can be made for this mass differences (and thus the oscillation frequency). The results are listed in Table 2.1 [23].

Table 2.1: Mass differences of neutral mesons from theory (calculating only the box diagram) and experiment. Units are time like, to relate to the oscillation frequency [23].

$\Delta M\left(\left\|K^0_{long} - K^0_{short}\right\|\right)$	
Theory	Experiment
$3.1 \pm 0.5 \cdot ns^{-1}$	$5.303 \pm 0.009 \cdot ns^{-1}$

$\Delta M\left(B^0_{d,heavy} - B^0_{d,light}\right)$		$\Delta M\left(B^0_{s,heavy} - B^0_{s,light}\right)$	
Theory	Experiment	Theory	Experiment
$0.46 \pm 0.2 \cdot ps^{-1}$	$0.472 \pm 0.017 \cdot ps^{-1}$	$15 \pm 5 \cdot ps^{-1}$	$> 14.5 \cdot ps^{-1}$ @ 95%CL

The deviations between theory and experiment in the kaon system are much larger than in the B-system due to so-called long distance effects. These effects originate from oscillations caused by final states common to both K^0 and $\overline{K^0}$, *e.g.* oscillations like $K^0 \rightarrow \pi\pi \rightarrow \overline{K^0}$. Note that these long distance effects are CKM suppressed for the B-system. Furthermore, the heavy top quark breaks the GIM symmetry between the diagrams relevant for the B system, which means that in this system the oscillation frequency is 'GIM enhanced'. Therefore, short distance effects dominate the $B\overline{B}$ oscillations, which significantly simplifies the theoretical description and makes it the preferred system for the analysis of processes with box diagrams. On top of that, the behavior of the heavy b-quark is in first order unaffected by QCD interactions with the light quark of the B-meson, a feature that is exploited with so-called heavy quark effective theories [24] to further reduce theoretical uncertainties.

The first successful measurements of a CKM-phase, involving a box diagram in the B-system, were presented by the B-factories BaBar and Belle. These experiments use an asymmetric e^+e^- collider (*i.e.* beams of unequal energy) with a center of mass energy equal to that of the $\Upsilon(4S)$ resonance, resulting in the production of coherent $B\overline{B}$ pairs with significant momenta in the lab frame [25]. They determined the value of $\sin(2\beta)$ (with $\beta = \arg\left(V_{td}V^*_{tb}/V_{cd}V^*_{cb}\right)$) to be 0.736 ± 0.049 based on the time dependent asymmetry[6] between the decay of $B^0_d \rightarrow J/\psi K_s$ and $\overline{B^0_d} \rightarrow J/\psi K_s$. These decays are also known as the

[6] An advantage of measuring a parameter via asymmetries is that some of the experimental uncertainties cancel.

'golden mode' due to their high branching ratio and signal to background ratio [19]. The interference between the two contributing amplitudes essentially originates in the same way as in the Gronau, London and Wyler analysis, since the K^0 and the $\overline{K^0}$ can form the same final state particle K^0_{short}. The two interfering Feynman diagrams at tree level are shown in Fig. 2.5. Note that the measurement contains the phase difference between a B_d $\left(\overline{B_d}\right)$ that decays directly into its final state and one that oscillates via the box diagram into a $\overline{B_d}$ (B_d) and then decays into the same final state. This phase difference is identified as 2β.

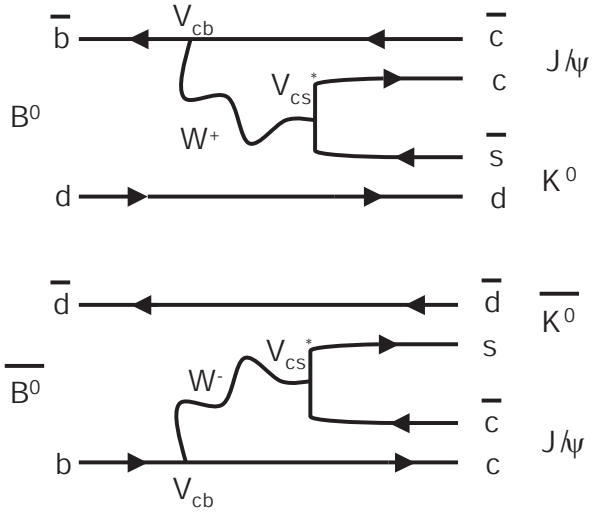


Figure 2.5: Feynman diagrams required for the calculation of $\sin(2\beta)$.

A similar measurement can be done with B_s-mesons. At the moment, such a measurement is considered even more interesting. First of all, because it allows to put a novel constraint on the unitarity triangles. Secondly, because the B_s life time difference appears to be much larger than expected, which could be a hint to a manifestation of new physics (although the errors on this measurement are large and more data are needed to be conclusive). The difference in decay width as measured by CDF[7] is $\Delta\Gamma/\Gamma = \left(65^{+25}_{-33} \pm 1\right)\%$, where the Standard Model predicts a difference of $7.4 \pm 2.4\%$ [26]. Note that the theoretical prediction for the difference in decay width between the $B^0_{d,heavy}$ and the $B^0_{d,light}$ is $0.24 \pm 0.06\%$, which is too small to be measured (current limit: $< 18\%$ @ 95% CL[8]).

2.4 Description of the $B_s \rightarrow J/\psi\phi$ decay

The 'golden mode' $B^0_d \rightarrow J/\psi K_s$ can be translated to the B_s system by changing the spectator quark from d to s. The interfering Feynman diagrams at tree level are shown in Fig. 2.6. Just as with the B_d-decay, the asymmetry is a measure for the phase difference between a decay with and without the box diagram. In this case, the phase difference is equal to $\arg\left(V_{ts}V^*_{tb}/V_{cs}V^*_{cb}\right)$, which can be associated[9] with the angle $\delta\gamma$ of Fig. 2.2. In the Standard

[7] See also http://www.ihep.ac.cn/data/ichep04/ppt/11_hq/11-0531-rescigno-m.pdf.

[8] See also http://www.slac.stanford.edu/xorg/hfag/results/index.html.

[9] Note that $\arg V_{tb} \approx \arg V_{us}$ and $\arg V_{cs} \approx \arg V_{cd}$.

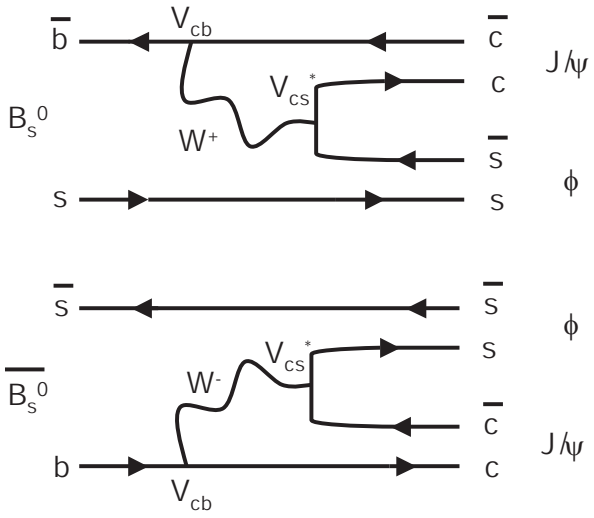


Figure 2.6: Feynman diagrams required for the calculation of $\delta\gamma$.

Model this angle is constrained to almost zero (by combining the knowledge of the phases of the other matrix elements with the unitarity conditions). The phase difference of the leading gluon interference diagrams (also known as penguin diagrams [27]) is equal to that of the tree level diagrams.

The cleanest measurement is provided for $B_s \to J/\psi\phi$ when the J/ψ subsequently decays into two muons and the ϕ into two charged kaons. Note that in the B_s system a time dependent asymmetry measurement is required, since the time integrated asymmetry is diluted due to fast oscillations. In contrast to the $B_d \to J/\psi K_s$, the B_s decays into two spin-1 particles. The angular distribution of the decay products of the spin-less B_s is isotropic in its rest frame. However, the subsequent decay of the spin 1 particles is not[10]. The angular distribution of the system can be described with

$$I\left(\theta_\phi,\phi_\phi,\theta_{J/\psi},\phi_{J/\psi}\right) = \Sigma_{\lambda=-1,0,1} D^{1}_{\lambda,0}{}^{*}\left(\theta_\phi,\phi_\phi,0\right) D^{1}_{\lambda,\pm1}{}^{*}\left(\theta_{J/\psi},\phi_{J/\psi},0\right) F^{J_{B_s}}_{\lambda\lambda}(t)\,. \qquad (2.9)$$

Where $D^{j}_{m'm}$ is the standard rotation matrix as given by Ref. [29], for each decay in its rest frame. The first common axis of the ϕ and the J/ψ (rest) frames is defined by their direction of motion (back to back) in the rest frame of the B_s. The angle θ_ϕ $\left(\theta_{J/\psi}\right)$is the opening angle of the kaons (muons) with respect to this axis, in the rest frame of the ϕ (J/ψ). The angles ϕ_ϕ and $\phi_{J/\psi}$ are not unambiguously defined. However, the difference between them, $\phi_\phi - \phi_{J/\psi}$, is the angular difference between the kaon and muon decay plane, as defined in the rest frame of the B_s (this angle does not change when the system is boosted to the rest frame of the ϕ or the J/ψ, since it is defined perpendicular to the direction of motion of these particles).

The first rotation matrix describes the angular distribution of $\phi \to K^+K^-$, with the three helicity states of the ϕ spin 1 particle and the spin-less kaons. The second component describes $J/\psi \to \mu^+\mu^-$, with three helicity states of the J/ψ spin 1 particle. The muons are

[10] As a result, the analysis of this decay is identical to $B^0_d \to J/\psi K^*$, which describes a spin 0 decay to two spin 1 particles as well (with the same constraints). This decay is discussed in detail in Ref. [28].

pair produced spin $\frac{1}{2}$ particles, which means that in first order their helicity is correlated[11], resulting in a projection of ± 1. Note that the spin projections of J/ψ and ϕ need to add up to 0 (to match the B_s spin projection), which is the reason that the sum only contains combinations with identical helicity states of these particles. Furthermore, the time dependence is completely contained in the B_s helicity amplitude $F_{\lambda\lambda}^{J_{Bs}}(t)$, which factorizes with the angular distributions. A complete discussion of this analysis is given in Refs. [30, 31].

The helicity amplitudes can be regrouped to match the CP eigenstates, where $A_{\parallel} = \frac{F_{++}+F_{--}}{\sqrt{2}}$ (CP even), $A_{\perp} = \frac{F_{++}+F_{--}}{\sqrt{2}}$ (CP odd) and $A_0 = F_{00}$ (CP even). In order to be able to regroup these components, transversity angles θ_{tr} and ϕ_{tr} can be defined as

$$\begin{aligned} \sin\theta_{tr}\cos\phi_{tr} &= \cos\theta_{J/\psi} \\ \sin\theta_{tr}\sin\phi_{tr} &= \sin\theta_{J/\psi}\cos\left(\phi_{\phi} - \phi_{J/\psi}\right) \\ \cos\theta_{tr} &= \sin\theta_{J/\psi}\sin\left(\phi_{\phi} - \phi_{J/\psi}\right). \end{aligned} \tag{2.10}$$

Note that the definition of the opening angle of the kaons (θ_{ϕ}) is the same as for the helicity frame. A graphical representation of the angles is given in Fig. 2.7. The decay plane in this so-called transversity frame is constructed from the trajectory of the two kaons in the rest frame of the B_s, which solves the ambiguous definition of this angle in the helicity frame. The decay geometry can now be described completely by the in-plane angle of the kaons (θ_{ϕ}) and the in-plane (θ_{tr}) and out-of-plane angle $(\phi_{tr})$of the muons.

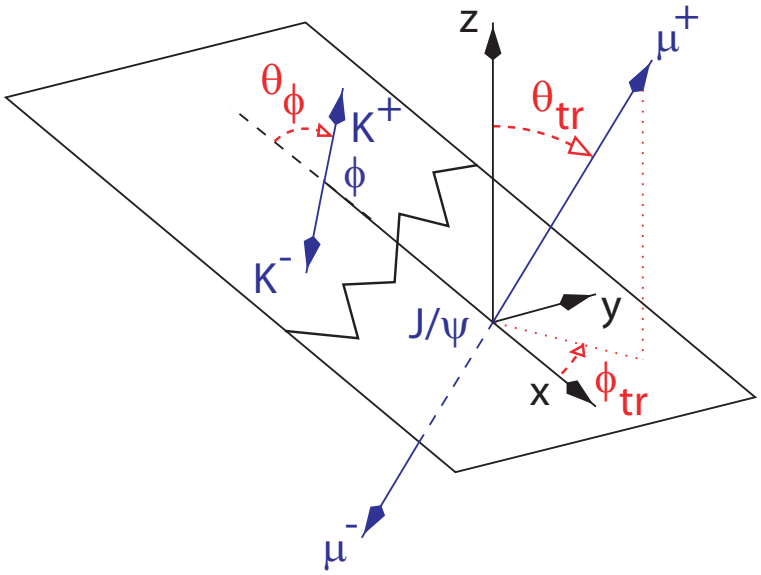


Figure 2.7: Geometrical representation of the decay $B_s \to (J\psi \to \mu\mu)(\phi \to K^+K^-)$ in the transversity frame, with a definition of the decay angles. Note that the decay plane is broken, to indicate that the angles of the ϕ and the J/ψ decays are defined in their distinct rest frames.

Inserting the $D^{j}_{m'm}$ functions and substituting the angular relations results in the following expression [32, 33].

[11] Note this correlation would have been perfect if muons were massless.

$$\begin{aligned}\frac{d^4\Gamma}{d\cos\theta_{tr}d\phi_{tr}d\cos\theta_\phi dt} \propto\ & 2|A_0|^2\left(1-\sin^2\theta_{tr}\cos^2\phi_{tr}\right)\cos^2\theta_\phi + |A_\perp|^2\sin^2\theta_{tr}\sin^2\theta_\phi \\ & + \left(\left|A_\parallel\right|^2\left(1-\sin^2\theta_{tr}\sin^2\phi_{tr}\right) - \Im\left(A_\parallel^* A_\perp\right)\sin 2\theta_{tr}\sin\phi_{tr}\right)\sin^2\theta_\phi \\ & + \frac{1}{\sqrt{2}}\left(\Re\left(A_\parallel^* A_0\right)\sin^2\theta_{tr}\sin 2\phi_{tr} + \Im\left(A_0^* A_\perp\right)\sin 2\theta_{tr}\cos\phi_{tr}\right)\sin 2\theta_\phi .\end{aligned} \tag{2.11}$$

Now, expression (2.11) can be integrated over θ_ϕ and ϕ_{tr}, without losing the information about the CP eigenstates of the components. Finally, one can group the CP even and CP odd contributions, $p(t) = |A_0|^2 + \left|A_\parallel\right|^2$ and $m(t) = |A_\perp|^2$, which results in

$$\frac{d^2\Gamma}{d\cos\theta_{tr}dt} = \frac{3}{8}p(t)\left(1+\cos^2\theta_{tr}\right) + \frac{3}{4}m(t)\sin^2\theta_{tr}. \tag{2.12}$$

The CP even state is associated with the $B_{s,light}$ and the CP odd state with the $B_{s,heavy}$ [34]. Their oscillating behavior is described by equation (2.8), which implies a time dependence of the contributions of $\cos^2\theta_{tr}$ and $\sin^2\theta_{tr}$ as indicated in equation (2.13) [35]. The subscripts in the equations indicate the particle types at $t = 0$, $p(0)$ and $m(0)$ are the decay rates at $t = 0$, $\overline{\Gamma} = \frac{\Gamma_{heavy}+\Gamma_{light}}{2}$ and $\Delta m = m_{B_s,heavy} - m_{B_s,light}$.

$$\begin{aligned} p_{B_s}(t) &= p(0)\left(e^{-\Gamma_{light}\cdot t} + e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right), \\ m_{B_s}(t) &= m(0)\left(e^{-\Gamma_{heavy}\cdot t} - e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right), \\ p_{\overline{B_s}}(t) &= p(0)\left(e^{-\Gamma_{light}\cdot t} - e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right), \\ m_{\overline{B_s}}(t) &= m(0)\left(e^{-\Gamma_{heavy}\cdot t} + e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right). \end{aligned} \tag{2.13}$$

These time dependencies clearly demonstrate that the contributions of $p(t)$ and $m(t)$ need to be disentangled to avoid dilution of the results. The partial contributions of $p(0)$ and $m(0)$ can be expressed as

$$R_T = \frac{\|A_\perp\|^2}{\|A_0\|^2 + \left\|A_\parallel\right\|^2 + \|A_\perp\|^2},$$

$$m(0) \propto R_T, \qquad p(0) \propto 1 - R_T, \qquad p(0) - m(0) \propto 1 - 2R_T. \tag{2.14}$$

This CP-odd fraction R_T can be determined from a time integrated measurement and will serve as an input for the time dependent analysis. The need to disentangle $p(t)$ and $m(t)$ increases when R_T is close to 0.5, as shown by equation (2.14). In chapter 6 the analysis of the decay $B_s \to J/\psi\phi$ is performed on simulated data. This analysis includes an optimization of the selection criteria, specific for this decay. The simulated data mimic the results of the *LHC*b experiment, which is currently under construction and is scheduled to become operational in 2007.

2.5 The current status of the unitarity triangle(s)

In section 2.2 and 2.3, two examples of recent measurements to determine the phases of the CKM triangles are mentioned. These results and other data, *e.g.* of the sides of the triangles

and of CP violation in the kaon system, are combined in a global fit [36]. The results are all consistent and the constraints on the first of the two unitarity triangles are shown in Fig. 2.8. Note that none of the B_s results mentioned in the previous sections are included in this fit and a direct measurement of the angle $\delta\gamma$, as described in the previous section, will probably not be available until *LHC*b is up and running.

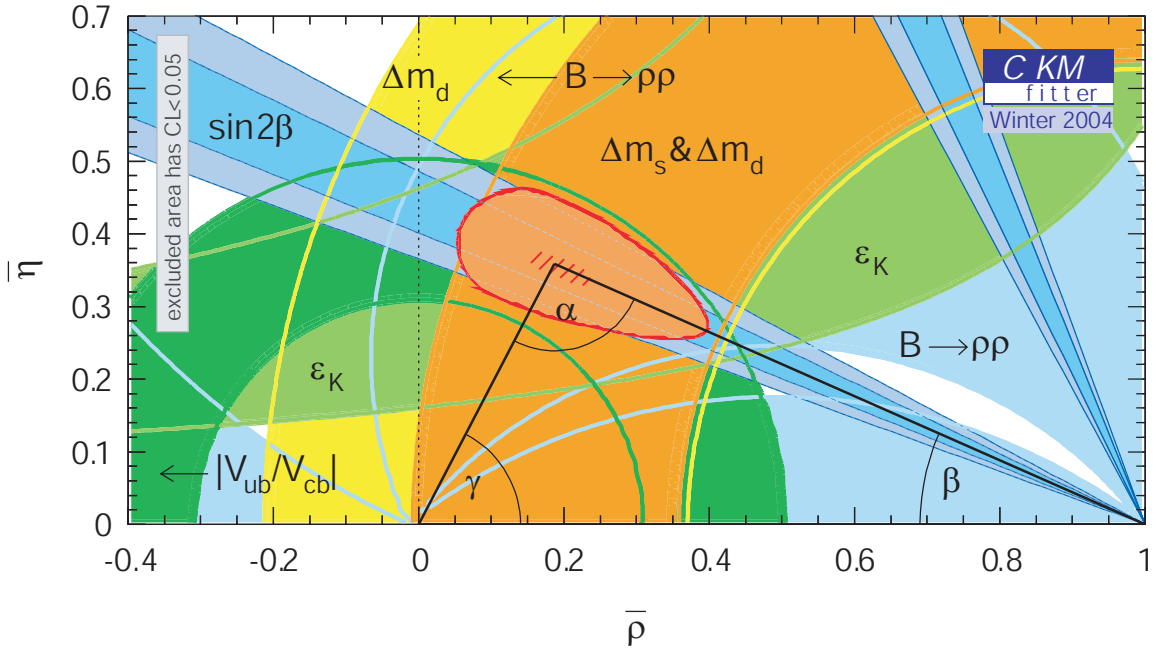


Figure 2.8: Current constraints on the first of the two CKM triangles. The angle β *is constraint by the measurements of the B-factories, as discussed in section 2.3. None of the recent* B_s *results are included in this plot.*

The constraints imposed by the $B_s \to J/\psi\phi$ analysis are part of the large B physics potential that *LHC*b offers to further constrain the unitarity triangles [1]. The goals of this experiment include direct measurements of all angles in various decay modes. On top of this, large statistics will offer the opportunity to look for signs of new physics in rare decays. A technical description of this experiment is given in chapter 3.

Chapter 3

Experimental setup

*LHC*b is an experiment designed to study CP violation and other rare phenomena in *B*-meson decays with high precision [1]. It will be installed in the large hadron collider, which is currently under construction and is expected to commence operation in 2007. The goal of *LHC*b is to check the theoretical predictions made in the standard model about quark mixing and to search for hints of new physics. In Table 3.1 the expected precision of the direct measurements after 1 year of data taking with the *LHC*b experiment are presented for the angles of the unitarity triangles. The present value of angle β is calculated directly from the measurements of the *B*-factories BaBar and Belle [37][1], the other 2 angles are results from a global fit of the CKM matrix using the standard model [38]. This global fit includes measurements of the lengths of the sides of the triangles, the angle β, direct *CP*-violation measurements and many others. Notice that the so-far unmeasured angle $\delta\gamma$ will be measured with the decay $B_s \to J/\psi\phi$ as discussed in chapter 6.

Table 3.1: Expected precision in the determination of the angles of the unitary triangles after 1 year of LHC*b data taking.*

Parameter	Decay mode	$value_{now}[rad]$	$\sigma_{LHCb}[rad]$
$\alpha = \pi - (\beta + \gamma)$	$B_d \to \pi^+\pi^-$	$1.6\pm0.2^*$	0.03
β	$B_d \to J/\psi K_s$	0.45 ± 0.03	0.03-0.16
$\gamma - 2\delta\gamma$	$B_s \to D_s^\pm K^\mp$	–	0.01
γ	any $B_d \to DK$	$1.1\pm0.2^*$	0.05-0.28
$\delta\gamma$	$B_s \to J/\psi\phi$	–	0.01
*Indirect measurements			

In this chapter is described what experimental equipment will be used for the measurements and how the various sub-systems cooperate to provide the results mentioned in Table 3.1. In section 3.1 a description is given of the large hadron collider and the production of heavy quarks. The *LHC*b detector is discussed in section 3.2, with emphasis on the excellent momentum mass and decay time resolution, as well as the particle identification features that make this detector optimal for these measurements. One of the sub-detectors, the vertex detector, is presented in more detail. This device plays a crucial role in the level-1 trigger.

[1] http://pdg.lbl.gov/pdg.html

The complications that arise from the tight integration of the vertex detector with the LHC beam vacuum are discussed in section 3.3.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring, superconducting circular accelerator with a circumference of 27 km. This project is currently under construction at CERN and aims to provide particle collisions with the maximum possible interaction energy to penetrate deep into the structure of matter.

The previous CERN accelerator was the Large Electron Positron (LEP) machine, which came into operation in 1989 and was shut down in 2001. Its electron - positron collisions were perfectly suited for detailed physics studies, since the electrons and positrons are 'point like' objects which can completely be converted into energy when they interact. However, circular machines of colliding electrons and positrons have reached their operational limits. Charged particle beams emit energy while they are forced to follow a circular path. This intense flux of synchrotron radiation drains the energy of the leptons and makes this kind of machine extremely expensive to operate much above 100 GeV per beam [39]. The emitted energy is proportional to the third power of the ratio of particle energy over its rest mass, which forms the basis of the idea to build a hadron accelerator in the same tunnel.

The design principles of all these accelerators are the same: particles are accelerated by electromagnetic cavities and held in a circular orbit by thousands of electromagnets. The proton's energy is, without a significant synchrotron radiation problem and with the technology available at the moment, limited only by the curvature of the ring and the electromagnetic field of the magnets. A machine like this can accelerate the particles to multi-TeV energies without unreasonable complications.

Figure 3.1 shows an overview of the main ring, with the 4 projected experiments: ATLAS, CMS, ALICE and *LHC*b. The smaller rings and the linear accelerators required for the injection of protons are shown as well.

Although the design principles of LHC are the same as for LEP, some major technical issues needed to be resolved. Two times eight superconducting cavities of niobium-coated copper provide the acceleration of the beams. The cavities operate at 4.5 K and are capable of delivering 2 MV each, with a field of 5 $MV \cdot m^{-1}$ at 400 MHz. The most critical requirement for these cavities is to keep the many proton bunches tightly spaced to ensure high luminosity at the collision points.

Furthermore, the current technology allows for mass production of superconducting magnets with a field strength of 8.33 T operating at 1.9 K. The acceleration of the proton beams is constrained to 7 TeV per beam [40] due to the field produced by 1296 of those dipoles. The resulting center of mass energy available for collisions of two beams circulating in opposite directions is $\sqrt{s} = 14\ TeV$. In Table 3.2 some of the LHC specifications are presented at its nominal design luminosity of $10^{34} cm^{-2} \cdot s^{-1}$.

Since the two LHC proton beams[2] will have identical charge, they must circulate through

[2] Anti-protons are available at CERN, but they are produced by a process with very low efficiency. Furthermore, at a collision energy of 14 TeV the dominating QCD interaction properties of protons and anti-protons are very similar.

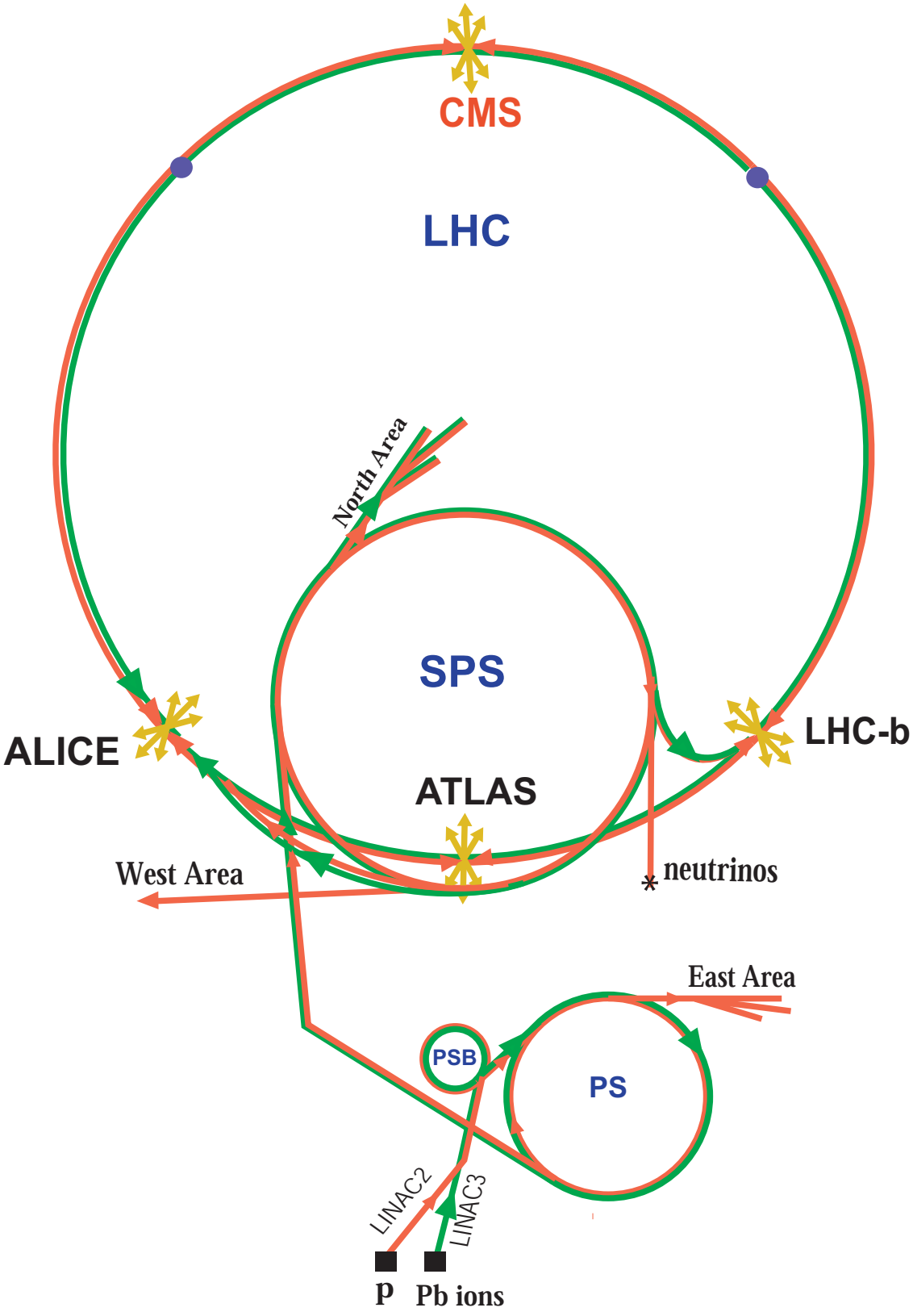


Figure 3.1: Schematic view of the CERN accelerators.

Table 3.2: Design specifications of LHC.

Center of mass collision energy $(\sqrt{s})$	14 TeV
Nominal design luminosity	$10^{34}\,cm^{-2} \cdot s^{-1}$
Collision frequency	40 MHz
Injection energy	450 GeV
Dipole field at 7 TeV beam energy	8.33 T
Beam current	0.56 A
Bunch spacing	7.5 m
Bunch separation	25.0 ns
Number of particles per bunch	$1 \cdot 10^{11}$
Energy loss per beam per turn	7 keV
Stored energy per beam	350 MJ
Lifetime	10 h

separate magnetic channels. There is no room for two separate rings of magnets in the tunnel, but a design with a twin-bore magnet which consists of two sets of coils and beam channels within the same mechanical structure and cryostat provides the solution. The two channels lie side by side in the cold yokes of the magnets. The penalty one must pay for this compact construction is that the magnets of the two channels cannot be aligned independently, nor can individual magnets with small errors in one channel be easily paired off with others having the opposite errors. However, LHC will be equipped with 510 lattice quadrupoles for optical tuning.

Large currents can be passed through superconductors of small cross section so that compact magnets can be built and operated at much lower costs than conventional magnets made from copper or aluminum. The only energy that is used in producing the field in the superconducting magnet is that for the refrigeration of the conductor necessary to retain its superconducting property. The heat capacity of the superconducting cables is very small at liquid helium temperatures, thus making these materials sensitive to any change in the deposition of energy on the cold surface of the magnet. The electromagnetic forces on the conductor increase with the square of the field strength and so does the stored electromagnetic energy. Thus, force-retaining structures and magnet quenching require special attention. A practical limitation in thermal stability comes from the total power delivered to the cold surface of the magnets by synchrotron radiation. Although this power is modest in comparison to an electron ring, it nevertheless poses a problem because it must be absorbed at a temperature of a few Kelvin.

The stainless steel vacuum chamber, in contact with the superconducting magnets, is protected from the power deposited by synchrotron radiation by a separate 'beam screen', cooled with its own cooling system. The standard pumping capacity of such a screen is insufficient to maintain the low operating pressure required to avoid beam blow-up[3].

[3]Synchrotron radiation from the proton beams creates photoelectrons at the beam screen wall. These photoelectrons are pulled toward the positively charged proton bunch. When they hit the opposite wall, they generate secondary electrons which can in turn be accelerated by the next bunch. This mechanism can lead to the fast build-up of an electron cloud, with potential implications for beam stability and heat load on the beam screen.

An experiment called COLDEX [41], designed and constructed by CERN in collaboration with NIKHEF, is used to study the performance of this vacuum system operating at cryogenic temperatures in the presence of a LHC type proton beam. With a proper design of the perforations in the 'beam screen', excessive gas can be transferred to the cold magnet surface, where the capacity is practically unlimited. The vacuum performance of this system is affected by various surface properties of the 'beam screen' like reflectivity, photo-electron yield and secondary electron yield. These properties are largely determined by the coating and the surface treatment. At the moment NEG (Non Evaporable Getter) coatings, the same materials as used in ion getter vacuum pumps, exhibit the most promising properties for stable vacuum operation. However, the secondary electron yield as well as other properties of activated and saturated NEG material are still under investigation.

3.1.1 Production of B-mesons

As discussed in chapter 2, analysis of B meson decays is interesting for the study of CP violation. In high energy pp-interactions the production mechanisms for heavy quarks are processes involving gluons and partons [42]. The relatively high mass of the b-quark allows for perturbative QCD calculations to make predictions about $B\overline{B}$ production and decay at LHC conditions.

From first order calculations it is already clear that LHC will be by far the most copious source of B mesons. The leading order Feynman diagrams of heavy-flavor production are typically categorized in three different types [43]. One of them, pair production, is shown in Fig. 3.2. The others are flavor excitation, with one gluon and one quark in the initial and final state ($Qg \rightarrow Qg$) and gluon splitting ($gg \rightarrow gg$, creating an additional $Q\overline{Q}$ pair in the parton shower).

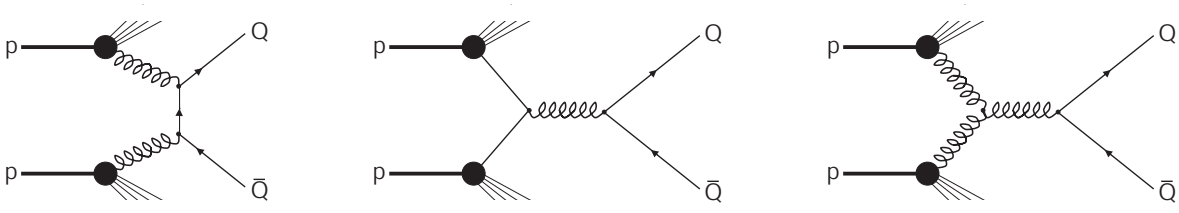


Figure 3.2: Feynman diagrams of the first order pair production mechanisms of heavy quarks at the LHC.

The production cross sections of these leading order processes are calculated by PYTHIA using the CTEQ4 particle distribution functions [44] (*i.e.* the proton structure functions as determined at HERA). This program includes elastic and diffractive proton interactions in order to obtain a realistic estimate of the background processes. The accuracy of these calculations is limited due to higher order corrections and non-perturbative hadronization effects. Estimates of these effects are included, based on a phenomenological (semi-empirical) parton shower model [45].

Both the b and the $\overline{b}$ recombine with other quarks during the hadronization process, which means that a variety of b hadron species will be produced. The broad longitudinal momentum distribution of the gluons in the protons (*i.e.* covering a wide range of Björken$-x$), combined

with their relatively low transverse momentum, results in the production of boosted b and $\overline{b}$ hadrons (each carrying one quark from the produced $b\overline{b}$ pair) along the beam axis in either the forward or the backward cone. This is demonstrated in Fig. 3.3 where the polar angles of the b and $\overline{b}$ hadrons calculated with the PYTHIA event generator are shown [1]. The average boost of these hadrons is about $\beta\gamma = 15$ and only few decay particles have momenta above 150 *GeV*.

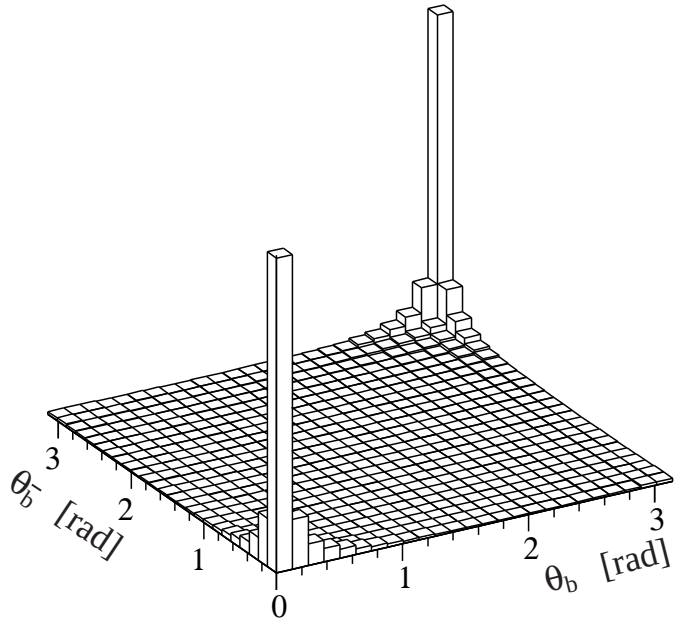


Figure 3.3: Polar angles of the b and $\overline{b}$ hadrons, calculated with the PYTHIA event generator.

3.2 An overview of *LHC*b

The availability of a large number of B mesons of various types and with many decay modes allows for measurements that provide new information about the CKM matrix elements. This information can be used to either over-constrain the unitarity triangles or cross-check results from different decay channels. Hence, a high precision B-physics experiment at LHC will provide the means to check the predictions made in the Standard Model about quark mixing and to search for evidence of new physics.

*LHC*b is specifically designed for this purpose, which makes the requirements for this detector different from those for the general purpose experiments. Here, some of the aspects are presented that make it feasible to perform detailed B-physics studies in a hadron collider with results that can significantly surpass those of the existing electron−positron colliders.

Issues related to B-meson production

Figure 3.4 gives an overview of the detector. *LHC*b does not require full 4π coverage, since the b and $\overline{b}$ hadrons are predominantly produced at small polar angles. The experiment is designed as a single arm spectrometer with a forward angular coverage from approximately

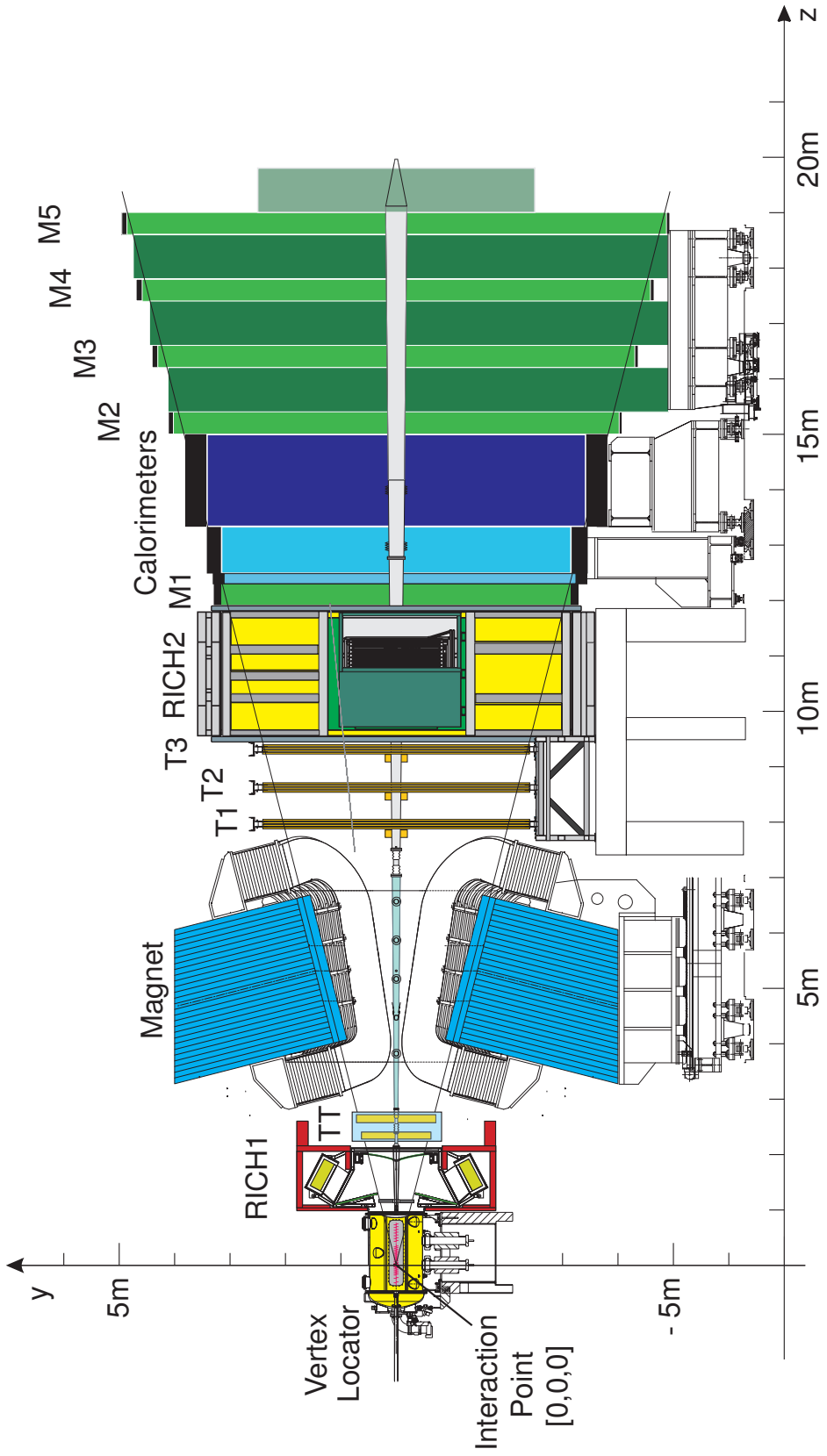


Figure 3.4: Schematic side view of the LHC*b experiment. The subdetectors are shown and labeled for reference. In the next paragraphs all these components will briefly be discussed.*

10 mrad to 300 mrad (250 mrad) in the bending (non-bending) plane [2]. This allows to reconstruct most of the B mesons in one of the two hemispheres.

In contrast to the general purpose detectors, ATLAS and CMS, this experiment will not run at the highest possible luminosity of 10^{34} cm^{-2}s^{-1}. Instead the beams will be defocused at the interaction point, to obtain the design luminosity of $2-5\cdot10^{32}$ cm^{-2}s^{-1}. This luminosity is chosen to limit the number of pp-interactions per event [46]. As a result, the events are relatively clean and easy to analyse in comparison to the events produced in the general purpose experiments. Furthermore, the detector occupancy is limited and the radiation damage is less severe.

The selected point of operation is indicated in Fig. 3.5, which shows the probability to have events with $0-5$ primary interactions as a function of luminosity. In combination with the assumed $b\bar{b}$ production cross section of about 500 μb the design luminosity results in an expected rate of 10^{12} $b\bar{b}$ pairs per year. Note that only 1 in about 160 interactions is expected to produce a $b\bar{b}$ pair.

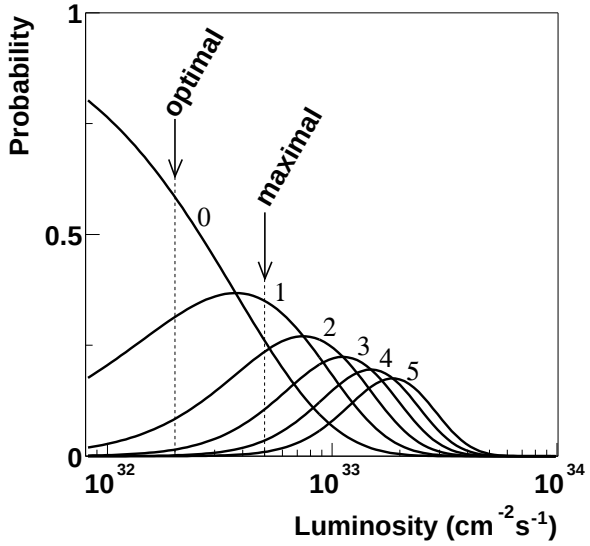


Figure 3.5: Probability distribution of the number of interactions per bunch crossing as function of the luminosity.

Issues related to B-meson decay

The main *LHC*b requirements for accurate identification and reconstruction of B decays can be associated directly to properties of the B mesons.

1. Events with B decays need to be distinguished from the rest with high efficiency. B mesons have a relatively high mass, which results in decay products with high transverse momenta. Furthermore, they have a long lifetime and they are produced at LHC with a high average boost. Therefore, they typically travel for several millimeters before they decay into other particles via the weak interaction. High transverse momenta of the decay products and the distance between the production and decay vertex are

the most important signatures for events with B decays and crucial components of the trigger, which is described in section 3.2.1.

2. The rapid oscillation of the B_s meson can only be resolved when the vertex resolution is sufficient. Of comparable importance is the momentum resolution, allowing accurate reconstruction of invariant masses and reduction of the vast background. The required vertex and momentum resolutions are obtained with a high performance tracking system and a magnet. Some details and the performance of this system are discussed in section 3.2.2.

3. B mesons have over a hundred different decay channels. The typical branching fractions of the interesting decays for CP measurements are of the order $10^{-5} - 10^{-6}$. Accurate distinction of all these channels is required to avoid systematic errors and dilution of the results, which means that particle identification, *e.g.* separating kaons from pions, is an absolute necessity for this experiment. In section 3.2.3 the basic design choices and performance of the particle identification detectors are discussed.

The detector response is studied with simulated events, which are produced by connecting the production software as discussed in section 3.1.1 with Geant4 to describe the weak decays and the interaction with the detector[4]. A dedicated package, called EvtGen, describes the decays of B hadrons[5], this program inter-operates with the Geant4 detector simulation, since the behavior of particles is affected by the materials they traverse. The results of these simulations served as input for the *LHC*b performance studies.

Each section is concluded with the appropriate performance numbers for the decay $B_s \rightarrow J/\psi\phi$. These numbers provide a direct link between the more technical description of the detector in this chapter and the analysis of simulated data as discussed in chapter 6.

3.2.1 Trigger

As mentioned in the previous section, only 1 in 160 events contains a B meson and the typical branching fractions of the interesting decays for CP measurements are of the order $10^{-5} - 10^{-6}$. Hence, an excellent trigger is required to ensure that only interesting events are selected for storage and off-line analysis. The trigger has three levels and is based on the identification of specific properties of $B\overline{B}$ production and decays.

Although this experiment runs at a lower luminosity than the general purpose experiments, there are still events with a high multiplicity and/or multiple interactions. At the moment it is investigated if *LHC*b is capable to reconstruct such events and separate multiple interactions. However, in the current configuration, the level-0 trigger rejects high-multiplicity events and thereby limits the processing time needed by the level-1 trigger. Another ingredient of the level-0 trigger is the selection of events containing decay products with a high transverse momentum, which is expected from the large B mass. The level-1 trigger exploits the large lifetime and high average boost of the B mesons with the identification of tracks that might originate from a decay vertex separated by a few millimeters from the production

[4] http://cern.ch/geant4

[5] See also http://www.slac.stanford.edu/~lange/EvtGen.

vertex. The higher level trigger uses as much of the tracking and particle identification as possible to find interesting B decays based on exclusive and inclusive selections.

Level-0 trigger

The purpose of the level-0 trigger is to reduce the data rate (initially equal to the LHC beam crossing rate of 40 MHz), to a rate at which in principal all sub-systems could be used for the derivation of a trigger decision. There are three sources of information at the disposal of the level-0 decision unit [47].

1. The pile-up system provides global event variables, such as an indication of the charged track multiplicity and the number of primary vertices of an event. The pile-up detector is located inside the vertex detector and provides digital hit information at a rate of 40 MHz. It consists of two silicon detector planes covering the full azimuthal angle at two different Z-positions, mounted in the 'backward direction', out of the acceptance of the rest of the detector. The sensors only measure the R-coordinate of the traversing particles. The pile-up trigger algorithm constructs tracks for each combination of hits, to identify the intersection with the beam trajectory (*i.e.* $R = 0$ in *LHC*b coordinates) as a possible primary vertex. Figure 3.6 shows a histogram constructed with this algorithm, where two primary vertices can be identified. When the first vertex is identified, the hits in the sensors that can be associated with the tracks pointing to this vertex are masked. Next, the same procedure is repeated with the remaining hits resulting in the hatched histogram, which shows a second peak. The height of this second peak is a measure for the number of tracks coming from a second vertex. The level-0 trigger rejects an event when the number of tracks coming from such a second vertex is above a certain threshold. The information from the pile-up detector can be used to select clean events, which reduces the probability that subsequent trigger levels erroneously select busy events without B mesons due to large combinatorics. Moreover, high multiplicity events would occupy a disproportional fraction of the data-flow bandwidth and processing power.

2. The calorimeter system is the second source of information for the level-0 trigger. It provides charged track multiplicity, an estimate of the total deposited transverse momentum and an estimate of the transverse momentum of the most energetic particles. The accuracy for electron, photon and π^0 candidates is $\frac{10\%}{\sqrt{E}} + 1\%$ and for the hadron candidates $\frac{80\%}{\sqrt{E}} + 10\%$ (where E is in GeV). Details about the components of the calorimeter and the 40 MHz processing for the level-0 trigger can be found in section 3.2.3.

3. The last source of information for the level-0 trigger is provided by the muon system. Stand-alone muon track reconstruction of this sub-detector provides a transverse momentum resolution of 20%. Dedicated processing units feed this information at a rate of 40 MHz to the level-0 trigger decision unit. The muon chambers are discussed in section 3.2.3.

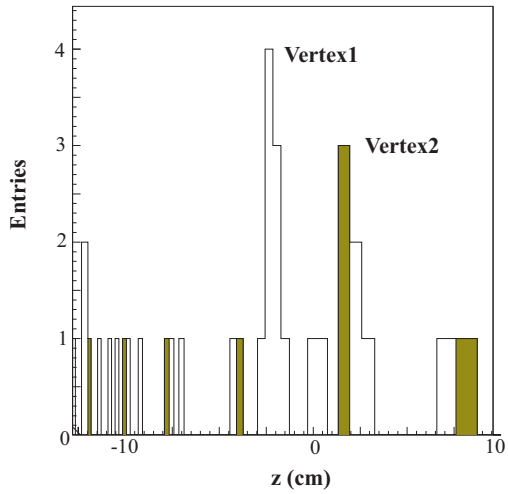


Figure 3.6: Histogram to determine the number of primary vertices. The first primary vertex is identified from the white histogram, its associated hits are masked and the hatched histogram is constructed from the remaining hits to see whether another primary vertex is present.

The level-0 decision unit collects all information from these components and is able to perform simple arithmetic to combine it into one decision per event. The thresholds of this trigger are shown in Table 3.3. The event rate provided to the level-1 trigger is reduced to 1 MHz with these settings. The efficiency of the level-0 trigger for muon channels is about 90% and for the remaining channels around 50%.

Table 3.3: Threshold of the level-0 trigger.

Pile-up system	
multiplicity 2^{nd} bunch crossing	multiplicity
< 3	< 112

Muon
$p_{T,\mu}$
$> 1.1\ GeV$

Calorimeter					
p_{T,e^-}	$p_{T,\gamma}$	p_{T,π^0}	$p_{T,hadron}$	Σp_T	multiplicity
$> 2.8\ GeV$	$> 2.6\ GeV$	$> 4.0\ GeV$	$> 3.6\ GeV$	$> 5.0\ GeV$	< 280

Level-1 trigger

The level-1 trigger is a software trigger that identifies events with a secondary vertex. It uses the hits in the vertex detector for stand alone track reconstruction in the *RZ*-plane as shown in Fig. 3.7. This implies that the vertex detector must be able to provide full event read-out at a rate of 1 MHz. The identification of possible *B* decay tracks is based on the impact parameter significance of the reconstructed tracks with respect to the closest primary vertex[6]. High impact parameter values indicate that tracks might originate from a secondary

[6]The primary vertices are determined from reconstructed tracks in the *RZ*-plane, using a similar method as for the pile-up system. An estimate of the XY position of these vertices is calculated as well, based on the division of the vertex detector *R*-sensors in 45° segments (which provides crude ϕ information).

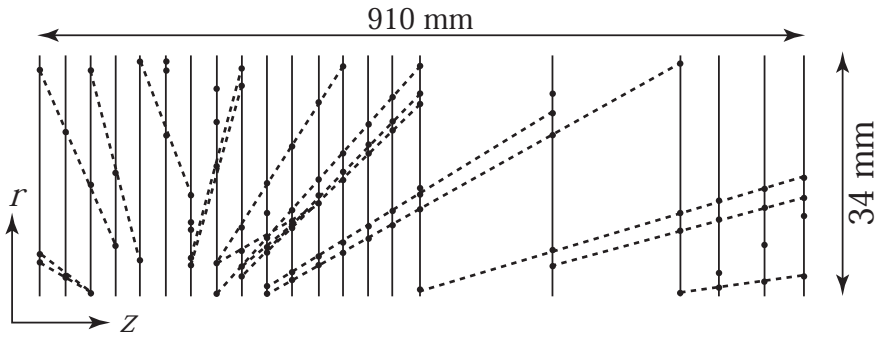


Figure 3.7: RZ *projection plot of hits in the vertex detector. Track reconstruction of these hits provides identification of high impact parameter tracks to the level-1 trigger.*

vertex. These tracks are reconstructed in three dimensions.

Tracks with a high impact parameter are matched to the muon or calorimeter clusters found by the level-0 trigger to make sure these particles have sufficient transverse momentum. If the match fails, then the signature of a possible secondary vertex is probably faked by multiple scattering of a low momentum particle and the candidate will be rejected. Simulations show that multiple scattering is one of the primary concerns for the overall trigger performance, especially when this scattering occurs before the particle traversed the first sensor. One of the most stringent requirements for the vertex detector is therefore to minimize the amount of material between the sensors and the interaction point.

The level-1 trigger uses the fringe field of the magnet between the vertex detector and the trigger tracker in combination with the reconstructed track segments in these sub-detectors. The curvature of the tracks in this region provides information about particle momenta with an accuracy of 20-40%. At the moment feasibility studies are performed to use information from the tracker stations behind the magnet in the level-1 trigger. This would significantly improve the momentum resolution at this trigger level.

The efficiency of the level-1 trigger varies between $45-80\%$ for the various channels. After the level-1 trigger, the event rate is reduced to 40 kHz. The remaining events contain particles with a high impact parameter significance, and a large transverse momentum, which are explicit signatures of B meson decay.

Higher level trigger

The higher level trigger (HLT) runs concurrently on the same CPU nodes as the level-1 trigger. The level-1 trigger takes priority due to its limited latency budget. The HLT algorithm starts with reconstructing tracks in the vertex detector and finding the location of the primary vertex. A fast pattern recognition program links the vertex tracks to information from the tracking stations. After reconfirmation of the level-1 trigger decision, the rate is reduced to about 20 kHz. Further selection is still under investigation, but is most likely based on full pattern recognition, including lepton identification but without the information from the ring imaging Cherenkov detectors (due to the processing time of the ring identification algorithm, see section 3.2.3). Two cases are distinguished:

1. Most of the important decay channels are identified via exclusive selection with re-

laxed selection criteria. In particular the side bands in the invariant mass distributions are necessary to be able to fit the contribution of the background. In addition, after this pre-selection it might even be possible to apply a partial RICH algorithm running at a few hundred Hertz to distinguish kaons from pions for these channels.

2. The rest of the channels will be identified with an inclusive selection. An example of such a selection would be $B_s \to J/\psi X$, which identifies any B_s-decay that includes a J/ψ.

The expected HLT efficiency is 95%, which results in a combined trigger efficiency for the decay $B_s \to J/\psi\phi$ of about 70% in case of $J/\psi \to \mu^+\mu^-$ and 60% in case of $J/\psi \to e^+e^-$. The output rate of the HLT is adjustable, but is expected to be 200 Hz at most. The annual amount of data that needs to be stored for off-line analysis is about 1 Petabyte. Some of the consequences of this enormous amount of data for the computing infrastructure are discussed in chapter 5.

3.2.2 Tracking detectors

The tracking system consists of a vertex detector, trigger tracker, magnet, inner tracker and outer tracker. First, the individual detectors will be briefly discussed and then an overview of tracking principles and combined performance are presented.

Introduction to the vertex detector

Figure 3.8 shows a cross section of the *LHC*b vertex detector. A complete description of this device is given in Ref. [48]. The design of the vertex detector has to comply with a variety of contradicting requirements. An overview of these requirements, together with the basic technology choices for this sub-detector is presented here. In section 3.3 the most critical components of the vertex detector are discussed in more detail.

The silicon sensor configuration is optimized to accurately identify particle trajectories with small polar angles. The innermost radius of these sensors should be as small as possible, since a short track extrapolation distance leads to a more precise impact parameter reconstruction. The proximity of these sensors to the interaction point implies that radiation hardness of sensors and read out electronics requires special attention.

The read out chip is called 'Beetle' and is specifically designed for the *LHC*b vertex detector, trigger tracker and inner tracker. It operates at 40 MHz and can deliver full analog event information at a rate of 1 MHz to the level-1 trigger for identification of tracks with a high impact parameter. A digital circuit provides real time binary read out to enable the implementation of the pile-up system in the level-0 trigger.

The shape of the output pulses, generated by the chip as a response to the signals received from the sensors, can be tuned with various parameters. An explanation of some of these parameters can be found in section 3.3.3, where the Beetle is discussed in more detail. The most interesting characteristics of the pulse shape are: rise time, signal/noise ratio and spill-over. Rise time is defined as the time difference between the moments where the pulse height amounts to 10% and 90% of the maximum pulse height. Spill-over is the value of the pulse height 25 ns after the pulse reached its maximum divided by the maximum pulse height.

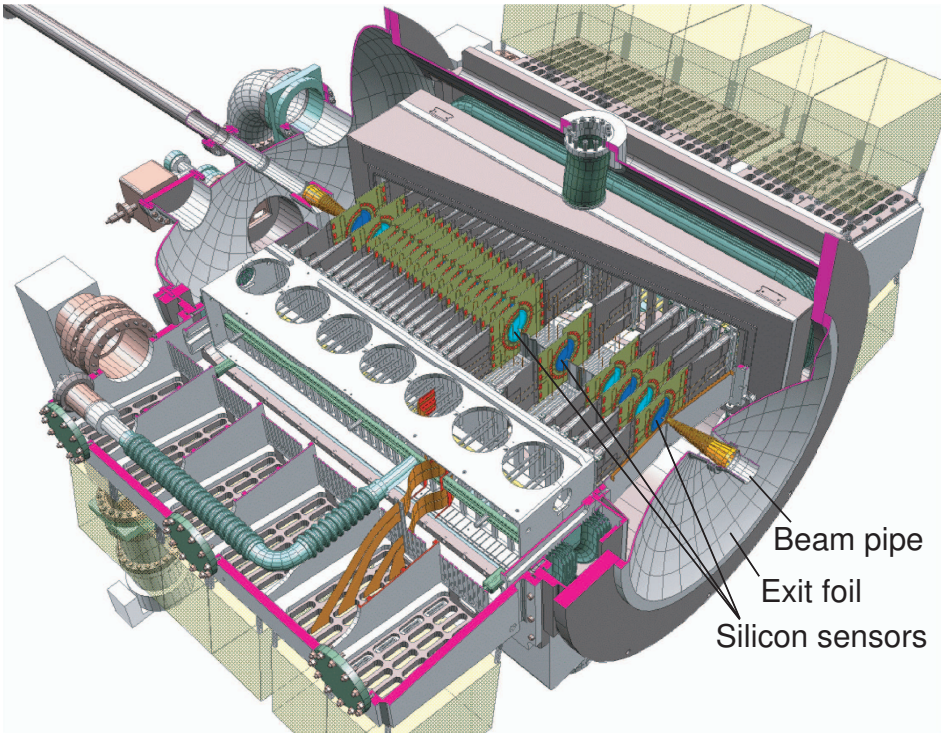


Figure 3.8: Cross section of the vertex detector. The beam pipe, silicon sensors and exit foil are indicated. This device is explained in more detail in section 3.3.

These two characteristics indicate whether the read-out is fast enough to process events that are only 25 ns, *i.e.* 1 bunch crossing, apart. Signal/noise ratio is the maximum pulse height divided by the width of the noise in the baseline.

Deflections of particles caused by multiple scattering, especially before they hit the first sensor, reduce the vertex resolution and might result in the identification of fake secondary vertices. Therefore, from a physics point of view, the best solution is to mount the silicon sensors directly in the beam vacuum. However, RF-coupling between the LHC beams and the vertex detector needs to be prevented as well as pollution of the beam vacuum by outgasing of detector materials and cables. This means a shield is required between the silicon sensors and the proton beams. The thickness of this shield and the thickness of the sensors themselves need to be minimized as much as possible. As a result the sensors are mounted with carbon fiber supports in a secondary vacuum, separated by a thin aluminum foil from the beam vacuum.

Trigger tracker

The trigger tracker, indicated with 'TT' in Fig. 3.4, is located just in front of the magnet. It consists of four 500 μm thick silicon strip layers. The first and the fourth layer have vertical read-out strips, the second and the third one have read-out strips rotated by a stereo angle of $+5^\circ$ and -5°, respectively. The four layers are arranged in two pairs, with a distance of 30 cm in between. The total active area of the detector is approximately 8.3 m^2 covered by 33 cm long strips with a pitch of 198 μm resulting in about 180,000 read-out channels. The occupancy varies between 0.3 and 3%. The trigger tracker is read-out by the same chip as the

vertex detector. The performance of the Beetle in the trigger tacker is less critical, since the sensors are thicker than those of the vertex detector. However, the larger strip length required special attention for the operation of this chip with large load capacitances connected to the front-end. Analysis of beam test measurements showed that the performance of this detector will largely exceed the most probable signal/noise ratio of 12.6 with a spill-over of 50% as used in the performance simulations.

The level-1 trigger uses data from this detector to assign transverse momentum information to tracks with a large impact parameter. Furthermore, in the off-line analysis its information is used to reconstruct the trajectories of long-lived neutral particles that decay outside the vertex detector and of low-momentum charged particles which are deflected out of the acceptance by the magnetic field before they can reach the inner or outer tracker.

Magnet

The magnet (see 'Magnet' in Fig. 3.4) is required to enable precise momentum measurements in the higher level trigger and off-line analysis. It consists of two trapezoidal coils bent at 45° on the two transverse sides, arranged inside an iron yoke. The magnet gap is wedge shaped in both vertical and horizontal planes, following the detector acceptance. The momentum and decay vertex resolution determine the accuracy of the reconstructed invariant mass of a particle and thus enable the suppression of background. The integrated field strength of the magnet is 4 Tm, large enough to ensure that the momentum resolution is completely dominated by multiple scattering effects (provided that the field is mapped accurately).

Outer and Inner tracker

Momenta of particles are measured in *LHC*b by reconstruction of track segments before and after the magnet and subsequent calculation of the track curvature in the magnetic field. The three tracking stations to determine the track segments behind the magnet are indicated with 'T1 − T3' in Fig. 3.4. As mentioned before, the momentum resolution is dominated by multiple scattering effects, hence every component in front of and in between the tracking stations has an extremely tight material budget. Straw drift-tubes are used for the outer parts of the tracking stations behind the magnet.

On the left side of Fig. 3.9 a front view of halve the detector is shown. The dimensions are about 2.5 m×5 m. Each tracking station contains 4×2 layers of straws as shown in the two most right hand pictures. A double layer of straws is called a module. The inner two of these modules have stereo angles of $+5^{\cdot\circ}$ and -5°. The tubes have a diameter of 5 mm. The wall of each tube consists of 40 μm kapton on the inside and 25 μm aluminum on the outside. The drift gas is a non-flammable mixture of Ar, CF_4 and CO_2, with a maximum drift time of 50 ns. Thus, a pulse generated by a traversing particle will also be seen during the next bunch crossing.

Cross-shaped arrays of silicon micro-strip sensors as shown in Fig. 3.10 cover the inner parts of the tracking stations behind the magnet. These so-called 'inner trackers' are installed to make sure that the average occupancy of the outer tracker stations stays below 7% per event. Each inner tracker has an area of about 0.3 m^2 and is constructed from 42 sensors with 320 μm thick single-sided p-on-n material. The strip length is 11 cm and the pitch is 198 μm, resulting in about 144,000 read-out channels with an average occupancy

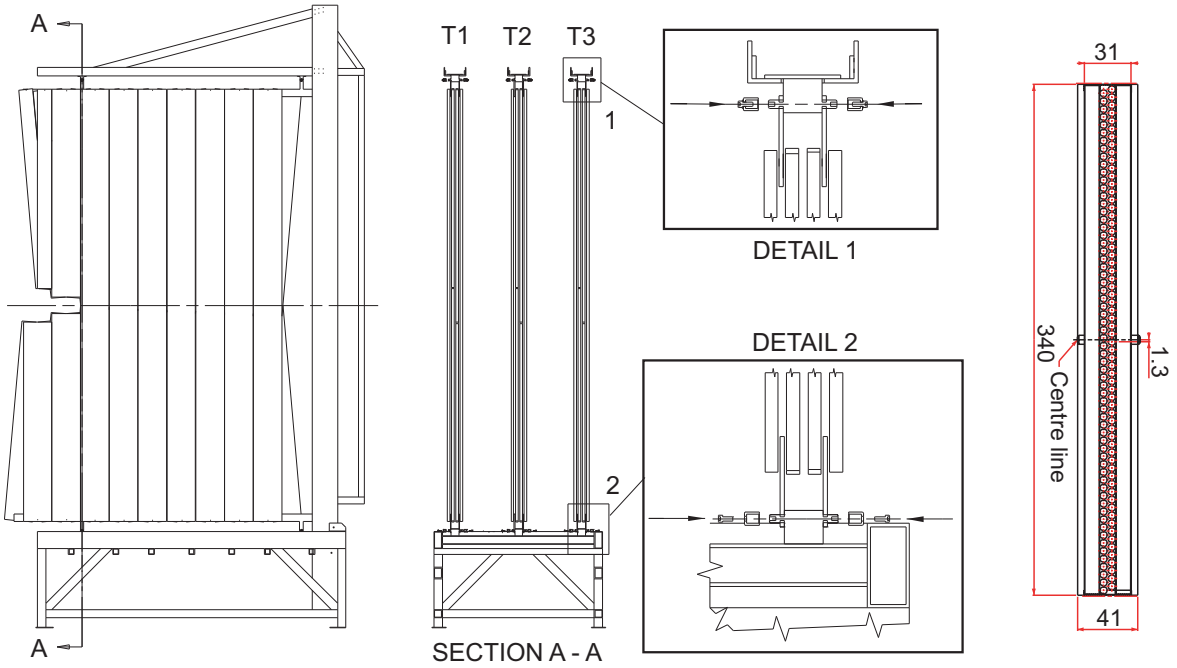


Figure 3.9: The picture on the left shows a front view of the outer tracker stations. The second picture shows a side view of the three stations. The next picture is a detail of one of the stations, showing the four modules that constitute such a station. The picture on the right shows a part of one of the modules. The individual straw tubes arranged in two layers are clearly visible. Dimensions are in mm

of about 0.5% per event. As with the trigger tracker, read out is done by the Beetle chip. The advantage of shorter strips is balanced with the choice for thinner detectors, making the requirements with respect to performance similar to those of the trigger tracker. A signal to noise ratio of 11 was measured with the Beetle1.1 for an input capacitance of 34 pF, which matches the longest strips. The equivalent noise charge (ENC) relation is measured to be $871e^- + 41.5e^- \times C\,[\mathrm{pF}]$. The selected pulse shape is almost symmetric with a full width halve maximum (FWHM) of 35 ns, corresponding to a rise time of about 25 ns and an estimated spill-over of 35%. Use of the Beetle1.3 will improve these numbers, although in principle they are sufficient for the inner tracker. At the moment feasibility studies are performed to use information from the inner and outer trackers in the level-1 trigger.

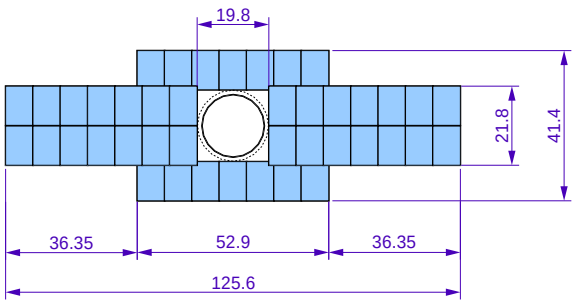


Figure 3.10: Cross-shaped silicon detector array of the inner tracker. Dimensions are in cm.

Tracking detectors combined

Figure 3.11 shows the various types of tracks, as identified by the track reconstruction process. Track reconstruction starts with the identification of tracks in the vertex detector. These so-called vertex tracks are used for a measurement of the primary vertex position and to identify tracks with high impact parameters.

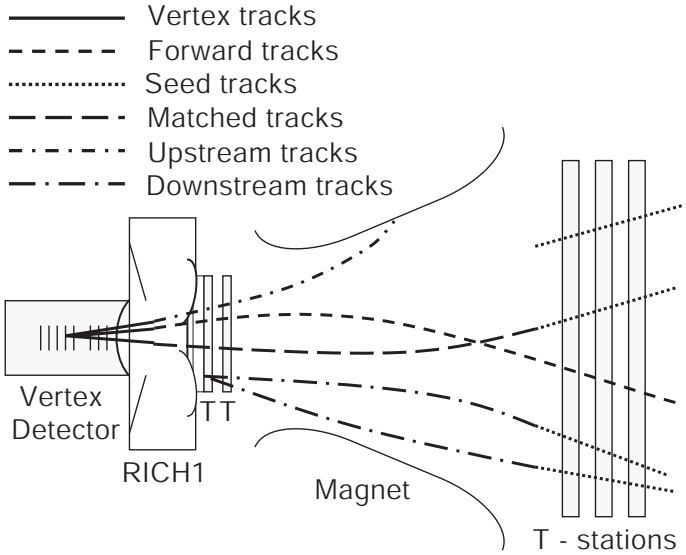


Figure 3.11: Different types of tracks.

The vertex tracks are propagated with a Kalman filter to the TT - stations before the magnet and the T - stations (inner and outer tracker) behind the magnet. About 95% of the *B*-decay tracks are identified with this technique and they are called 'forward tracks'. The remaining hits in the T - stations are used to create 'seed tracks', which play an important role in the RICH2 particle identification algorithm.

Seed tracks that can be matched with remaining tracks in the vertex detector are called 'matched tracks', they account for the identification of an extra 2% of *B*-decay tracks. This algorithm is more robust than the Kalman filter, but can only be applied to a small number of remaining tracks to avoid large combinatorics. The matched tracks and forward tracks together are referred to as 'long tracks' and are the most interesting for physics analysis. These tracks traverse all sub-detectors and allow for proper reconstruction of the complete decay.

The low momentum charged particles will not reach the T-stations due to deflection by the magnet; they are called 'Upstream tracks'. Another type of tracks are the 'Downstream tracks', *e.g.* decay products of K_s particles. Of these particles 25% decays in the vertex detector and will be identified as long tracks, 50% decays outside the vertex detector but before the magnet and will be identified as 'Downstream tracks', the rest of the K_s particles will not be reconstructed.

The *LHC*b performance simulations show that the combined information of the tracking detectors will allow to reconstruct the decay length of a *B*-meson with a resolution of about 22 μm in the transverse and 144 μm in the longitudinal direction. In combination with an average momentum resolution of 0.4%, this means we can resolve B_s-oscillations with 5 σ

sensitivity up to a mass difference of 68 ps^{-1} within 1 year of data taking, *e.g.* with $B_s \rightarrow D_s\pi$. The Standard Model prediction of this mass difference is 20 - 25 ps^{-1}, but new physics may drastically affect this value.

3.2.3 Particle identification detectors

Ring imaging Cherenkov detector

The most important purpose of the ring imaging Cherenkov (RICH) detector is separation of charged kaons and pions. The need for this separation is demonstrated in Fig 3.12. Each decay channel shown in this figure will provide interesting information if it can be distinguished from the others (*e.g.* $B_s \rightarrow D_s\pi$ provides a measurement of the miss-tag rate and of Δm_s, while $B_s \rightarrow D_sK$ measures $\gamma - 2\cdot\delta\gamma$). Particle identification is required in order to unravel these decays, since their topology is identical.

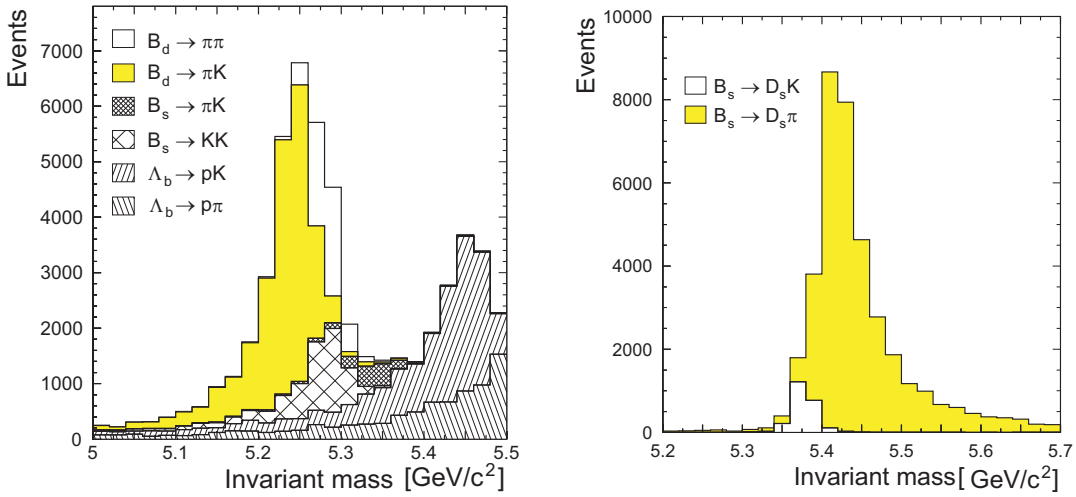


Figure 3.12: Simulation showing the overlapping invariant masses of kaon and pion decay modes.

The RICH offers identification of charged particles by detection of Cherenkov light produced by these particles when they travel through various radiators. The circular distribution of the Cherenkov light can be associated with a kaon or pion traveling in a particular direction with a well-determined momentum. It is not possible to use this detector in the level-0 and level-1 trigger, since the particle identification algorithm needs the trajectory and the momentum of the particle. At the moment a fast algorithm is under development to process information of the RICH detectors for the higher level trigger. Three kinds of radiators with different refractive indices are required, since there is a strong correlation between the polar angle of the emitted Cherenkov light and particle momentum[7]. For the lowest momentum particles (up to about 10 GeV) 'Silica aerogel' is used, for the intermediate region (up to about 50 GeV) gaseous C_4F_{10} and for the particles with the highest momentum (up to about 150 GeV) gaseous CF_4.

[7]

Material	Silica aerogel	C_4F_{10}	CF_4
Refractive index (η-1)	$3.2\cdot10^{-2}$	$1.4\cdot10^{-3}$	$4.7\cdot10^{-4}$

The low and intermediate momentum charged particles are identified by RICH1, which is located before the trigger tracker and the magnet (see Fig. 3.4) to limit the detection area and to be able to also assign particle types to the 'VTT tracks', mentioned in the previous section. The layout of the detector is shown in Fig. 3.13. The need of an entrance window is avoided by mounting the detector directly onto the exit window of the vertex detector vessel. Two slightly tilted spherical mirrors of either a carbon-composite or beryllium will project the Cherenkov light via two flat mirrors onto pixel hybrid photon detectors (HPD's). Both the flat mirrors and the HPD's are mounted outside the acceptance of the experiment.

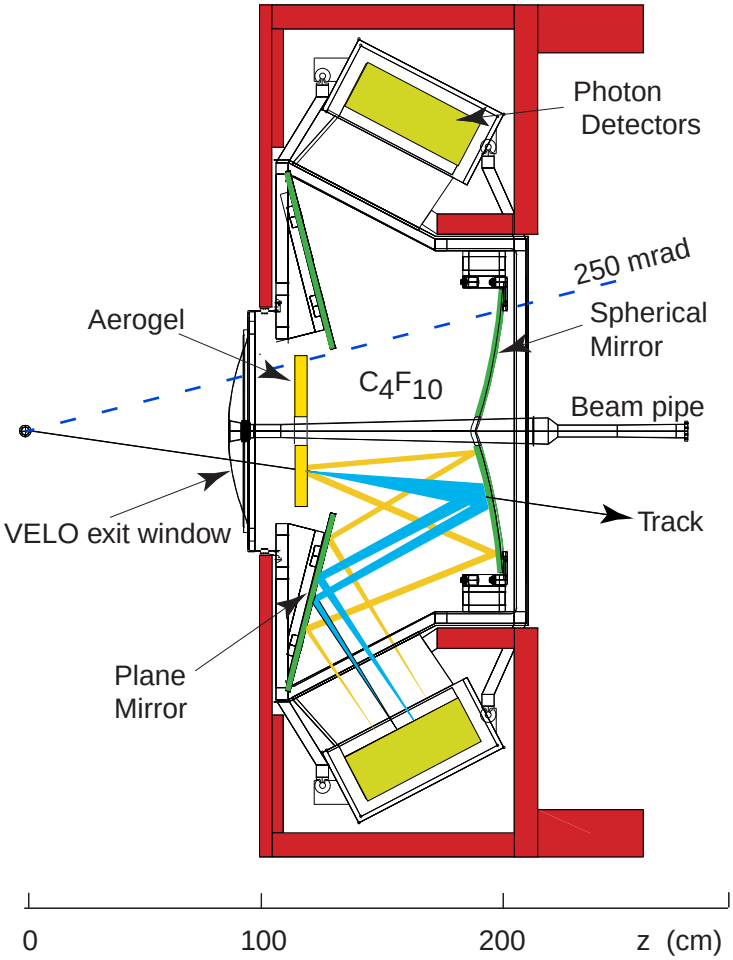


Figure 3.13: Layout of the RICH1 detector.

The high resolution HPDs are specifically designed for *LHC*b. They are sensitive for visible light down to wavelengths of 200 nm . Their anodes consist of a silicon sensors containing 256×32 pixels, bump bonded to 40 MHz pixel read-out chips, encapsulated inside the vacuum envelope. The photoelectrons, released by a single photon incident on the photo-cathode are accelerated onto the silicon sensor by a high voltage of 20 kV, resulting in a signal of about 5,000 electrons in the silicon. The expected equivalent noise charge of these detectors is below 300 electrons, thus there will be no problem to reach the 85% detection efficiency as assumed in the performance simulations. Although distortions in the electron optics are large due to the presence of magnetic fringe fields up to 2.5 mT, tests confirmed

that the HPDs can properly be operated when corrections are applied for these effects. The maximum expected radiation dose for these detectors is about 3 kRad $\cdot$ year^{-1}, which should be no problem for the employed silicon detectors.

The high momentum charged particles are identified in RICH2, which is located between the tracking stations and the calorimeters. The material budget is not as tight as for RICH1, but electrons and hadrons should be able to traverse the detector and reach the calorimeters. Two spherical mirrors of polished 6 mm thick glass cover a region up to 120 (bending plane)$\times$100 (non$-$bending plane)mrad^2 around the beam pipe to reflect the Cherenkov light emitted by high momentum particles traveling in the most forward direction. High momentum particles with sufficient transverse momentum to escape outside the acceptance window of the RICH2 detector are most likely not related to B-meson decays. The mirrors are slightly tilted and the light is directed via flat mirrors onto HPDs mounted outside the acceptance, just like in RICH1. The magnetic fringe field as well as the radiation level are lower for RICH2 than for RICH1, thus the same type of equipment can be used without further complications.

Calorimeters

The calorimeters, indicated with 'Calorimeters' in Fig. 3.4, provide information to the level-0 trigger about high transverse energy depositions possibly caused by hadrons, electrons or photons. Their most demanding task is the identification of electrons, which needs to reject 99% of the inelastic pp interactions. Performance simulations show that this provides an enrichment of at least a factor 15 in B events. Furthermore, the calorimeters identify π^0s and prompt photons in addition to the kaon and pion identification offered by the RICH detectors. The calorimeter system, presented in Fig. 3.14 contains several subsystems.

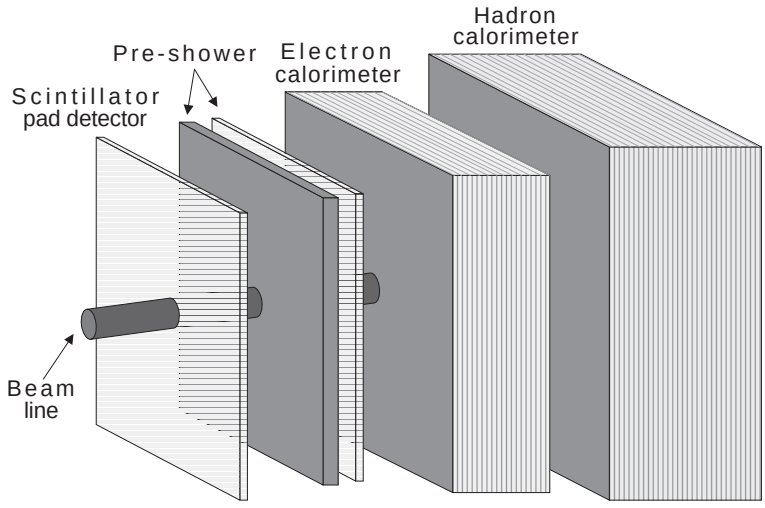


Figure 3.14: Exploded view of the calorimeter system.

The scintillator pad detector (SPD) uses about 6,000 scintillator pads with a surface of 7.6 m$\times$6.2 m to detect charged particles before they enter the calorimeter. After the SPD a pre-shower (PS) system, which consists of a 12 mm thick lead converter and again a layer of

about 6,000 scintillator pads, identifies the start of electron and photon showers. The electromagnetic showers will continue in the electromagnetic calorimeter (ECAL). The ECAL uses a shaslik made from 66 layers with 2 mm thick lead absorber plates interspaced with 4 mm thick scintillating tiles, resulting in a total length of 25 X_0. This sub-system measures the remainder of the electromagnetic showers as well as the start of hadronic showers. The hadrons will not be absorbed completely in the ECAL and continue their trajectory in the Hadron Calorimeter (HCAL). The HCAL is a sampling device made out of steel and scintillating tiles mainly used for level-0 trigger purposes and for the identification of long-living neutral hadrons. Furthermore, it prevents everything but muons from entering the muon chambers.

The radiation levels are the highest in the ECAL and HCAL, where electronic and hadronic showers are present. The regions closest to the beam pipe will experience about 25 krad (ECAL) and 50 krad (HCAL) a year. Extensive calculations and tests were performed to determine the effects of such a dose on the performance of the calorimeter system. A conservative estimate for 10 years of LHC operation shows an increase from 0.8% to 1.5% in the constant term of the momentum resolution for the ECAL and an increase from 10% to 12% for the HCAL. Based on these numbers no substantial performance degradation is expected within 10 years of LHC operation.

The scintillators are connected to photomultiplier tubes of the same type for all the calorimeter sub-systems. The gain of the tubes is adjusted, to make them sensitive to about 25 photons in the SPD and PS system resulting from pions of a few MeV, where in the ECAL and HCAL the deposited energy of high energy particles can reach 200 GeV, resulting in up to 200,000 photons. The use of one type of photomultiplier tubes means that the signal shapes will be about the same in all calorimeter sub-systems. The pulses are short enough to be able to separate signals which are 25 ns apart. The difference in signal height has some implications for the front-end system, where the signals are digitized by a 10 bits ADC in case of the SPD and PS system, but a 12 bits ADC is selected for the extended range of the ECAL and HCAL. As a result, the pedestal stability is less important for SPD and PS system. On the other hand, the smaller signals in the SPD and the PS system result in large fluctuations and the signal should be integrated within the 25 ns window of an event to give stable results.

The information of the SPD, PS, ECAL and HCAL is used for the level-0 trigger decision. This means the front-end electronics needs to provide a 40 MHz read-out frequency, which constitutes the most difficult requirement to meet for the calorimeter system.

The adopted solution for the SPD and the PS involves the alternation between two integrators every 25 ns. One integrator is reset when the other is active. The front-end system of the SPD converts the signal to a simple discriminator output to confirm the detection of charged particles. The front-end electronics of the PS system provides the full analog signal to allow for a pulse height analysis that measures the energy content of the first part of the showers. The PS information for the level-0 trigger is pedestal and gain corrected and converted to an 8 bits value via a look-up table.

The electronic systems for the ECAL and HCAL are identical due to their common functionality and the similarity of the input signals. One front-end board handles 32 scintillator cells. The signals of these cells are converted to 12 bits data and processed by FPGAs. The individual signals are stored in 4 μs (160 cells) fixed latency buffers, awaiting the level-0

decision. The highest transverse energy deposition of any 2×2 cell combination is selected on each front-end board for level-0 trigger processing. The signals of cell combinations that span multiple front-end boards are routed via a dedicated backplane or, in a few cases, via cables. The level-0 trigger processing continues in a dedicated crate that selects the best candidates for electron, photon, π^0 and hadron with the highest transverse energy of all front-end boards. This information, together with the information about multiplicity and total deposited transverse energy in the calorimeters is sent to the level-0 trigger decision unit.

Muon detector

The position of the muon chambers behind all other *LHC*b sub-detectors (see Fig. 3.4) is based on the penetrative power of muons. The exception is M1, which is located in front of the calorimeter because multiple scattering in the calorimeter would reduce the muon track resolution. This first muon chamber could in principle be switched with the SPD to reduce the amount of material and thus the amount of photon conversions in front of the SPD. However, shadows from small angle deflections (harming the level-0 trigger processing of the calorimeters) are reduced by putting the SPD as close as possible to the ECAL and by exact matching of the projective pad layout of both detectors. The thickness of M1 is tuned to optimize the π^0 reconstruction of the calorimeter whilst sacrificing as little muon track resolution as possible.

Multi-Wire Proportional Chambers (MWPCs) as shown in Fig. 3.15 are selected as detector technology, primarily based on rate capability and time resolution. The chambers have symmetric cells with an anode-cathode gap of 2.5 mm and an anode-wire spacing of 1.5 mm. Wires are grouped in pads of 4 to 42 to match the required granularity, varying from 6 mm to 62 mm. The MWPC gas is a non-flammable mixture of Ar, CO_2 and CF_4, with a low sensitivity to neutron background. Iron absorbers with a thickness of 80 cm are located in between the chambers (except for M1 and M2). The total detector area is about 435 m^2.

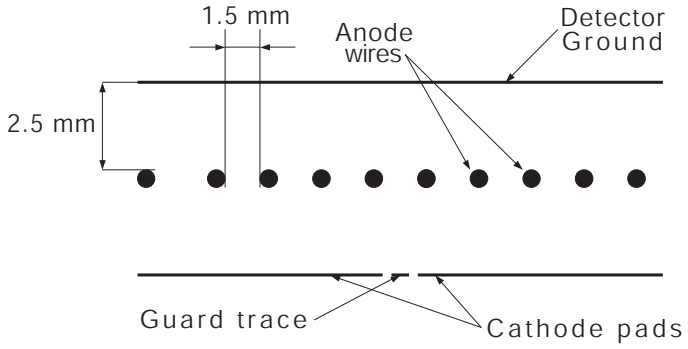


Figure 3.15: Schematic diagram of a Multi-Wire Proportional chamber as used in the muon detector.

The primary task of the muon system is the identification of high momentum muons for the level-0 trigger. Each muon chamber is divided in 4 regions (see Fig. 3.16) and each region is divided in 12 sub-regions for level-0 trigger processing. Dedicated processing units combine the information from the same sub-region of all five muon stations, forming a

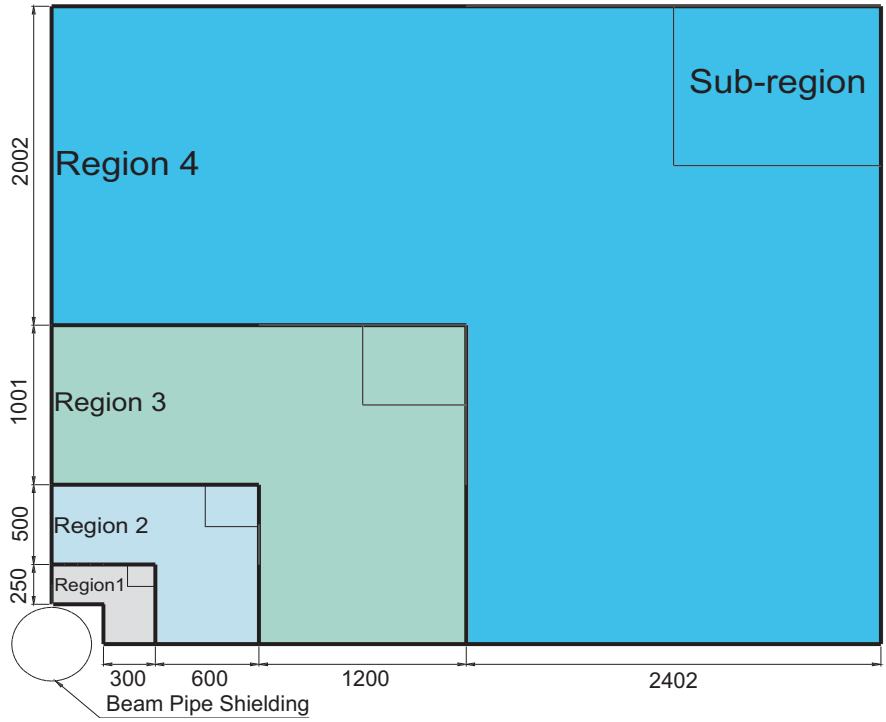


Figure 3.16: Front view of one quadrant of muon station 2, showing the dimensions of the regions in mm. In the upper right corner of each region, one of its 12 sub-regions is indicated.

tower pointing toward the interaction point. The processing unit of one tower also receives information from neighboring towers. Hits in M3 function as track seed and confirmation is obtained by a straight line extrapolation from the assumed interaction point (*i.e.* $[0, 0, 0]$ in *LHC*b coordinates) to stations M2, M4 and M5. For all three stations a hit must be found within a specified region from the intersection of this line with the detector plane to positively identify a muon track. Just like in the calorimeters, the logical layout is projective. There is a one-to-one mapping from pads in M3 to M2, M4 and M5, as well as from pairs of pads in M2 and M3 to M1. This allows the track-finding algorithm to be implemented using only logical operations. Once track finding is completed, the transverse momentum is determined with look-up tables with the hits in M1 and M2. Up to eight selected tracks with the highest transverse momentum are sent to the level-0 trigger decision unit.

The other important task of the muon system is efficient muon identification, both for reconstruction of physics channels with muons in the final state as well as for tagging purposes. The minimum momentum for a muon to penetrate to M3 is 3 GeV. The requirements for a positive muon identification depend on the momentum. If the particle has a momentum of less than 6 GeV, it is identified as a muon when there are hits in the selected areas of interest of M2 and M3. If the particle has a momentum of more than 6 GeV, but less than 10 GeV, hits are required in the selected areas of interest of M2, M3 and either M4 or M5. For particles with momenta above 10 GeV, M2 - M5 should all have hits in the selected areas of interest. The selected areas of interest are based on a parametrization as a function of momentum in the 4 regions in each station. Physics studies show that the muon identification efficiency is about 94%. The misidentification probabilities for electrons, pions, kaons and protons is around 2%.

Particle identification detectors combined

The combined particle identification algorithms are based on estimators from the RICH, calorimeters and muon system. The estimators express the ratio of the likelihood between the particle type and background hypotheses (*e.g.* the muon system tests the hypotheses of a particle being a muon versus not being a muon). The information of the various sub-detectors can be combined by multiplication of the estimators of each sub-detector for a certain particle type. The weighting of these various sources of information is not at all trivial and the optimal solution partially depends on the subsequent analysis. A more detailed discussion of this problem can be found in Ref. [2].

The hypotheses of two particle types can be combined by subtraction of the two log-likelihood values, which provides the log of the ratio of one particle type hypothesis over the other. Simulated log-likelihood distributions for various particles are shown in Fig. 3.17.

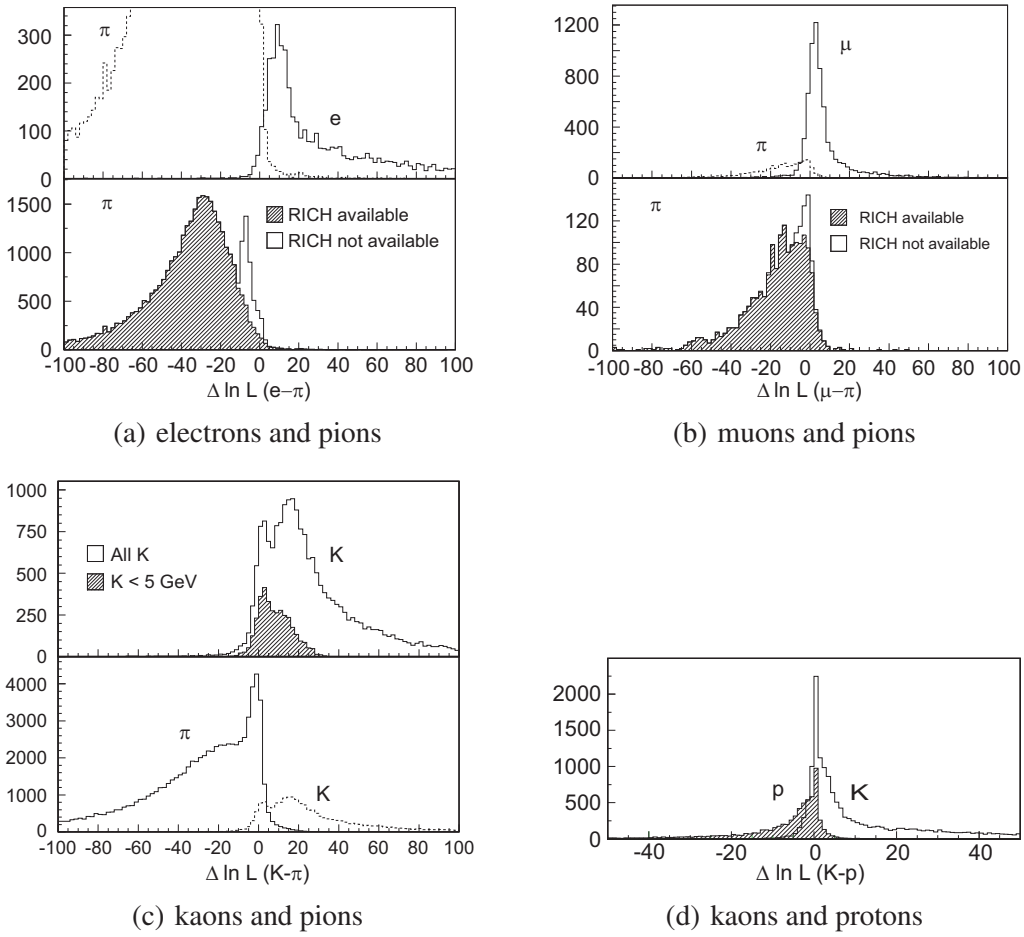
The plots with electrons or muons versus pions (plot a and b) are divided in a top and a bottom part. The bottom part shows the log-likelihood distributions of the pions without RICH information (*i.e.* for the level-1 trigger) and with RICH information (*i.e.* in case of higher level trigger or off-line analysis). The top part shows the pion log-likelihood distribution with RICH information and the electron or muon log-likelihood distribution. The plot with kaons versus pions (plot c) is divided in a top and a bottom part as well. In this case the top part shows the log-likelihood distribution of all kaons and of kaons with a momentum below 5 GeV. The bottom part shows the log-likelihood distributions of pions and of all kaons. The plot with kaons versus protons (plot d) shows the log-likelihood distributions of protons and kaons.

The log-likelihood values for separation of electrons, pions, muons, kaons and protons can be optimized to achieve the best signal to background ratio. Details about this optimization for the $B_s \rightarrow J/\psi\phi$ decay can be found in chapter 6. The selected log-likelihood value for muon/pion separation is -2, resulting in 90% muon efficiency and 1% pion misidentification. For kaon/pion separation a value of -3 is selected, resulting in 88% kaon efficiency and 3% pion misidentification.

3.3 The vertex detector in more detail

The central issue in the design and operation of the vertex detector is a compromise between two requirements. On the one hand, from the LHC machine point of view, the four experiments should make sure they never seriously hamper operation of the accelerator and therefore each other. On the other hand, from the *LHC*b physics point of view, material between the sensors and the proton interactions should be avoided as much as possible.

A solution for this problem is to install the silicon micro-strip sensors of the vertex detector in a secondary vacuum close to the beams. As a consequence the *LHC*b vertex detector is tightly integrated with the beam vacuum system. The impact of this tight integration on the design and operation of the vertex detector is the main topic of section 3.3.1. Next, the silicon sensors are described in section 3.3.2, especially the complications that arise from their proximity to the interaction point, followed by a discussion of their cooling system. The chapter ends with a description of the radiation hard analog read-out chip called 'Beetle' in section 3.3.3. Beam-test results from an experiment with this chip is discussed in chapter 4.



(a) electrons and pions

(b) muons and pions

(c) kaons and pions

(d) kaons and protons

*Figure 3.17: Particle identification using the log-likelihood difference. The plots with electrons or muons versus pions (plot a and b) are divided in a top and a bottom part. The bottom part shows the log-likelihood distributions of the pions without RICH information (*i.e. *for the level-1 trigger) and with RICH information (*i.e. *in case of higher level trigger or off-line analysis). The top part shows the pion log-likelihood distribution with RICH information and the electron or muon log-likelihood distribution. The plot with kaons versus pions (plot c) is divided in a top and a bottom part as well. In this case the top part shows the log-likelihood distribution of all kaons and of kaons with a momentum below 5 GeV. The bottom part shows the log-likelihood distributions of pions and of all kaons. The plot with kaons versus protons (plot d) shows the log-likelihood distributions of protons and kaons.*

3.3.1 Vertex detector design

Schematically, the silicon strip sensors are arranged as shown in Fig. 3.18 [48]. The silicon sensors are positioned such that particle trajectories with small polar angles (thus within the detector acceptance) are almost perpendicular to the detector planes and traverse at least three stations, which means they can be reconstructed accurately.

The active area of the silicon sensors starts at 8 mm from the beams. During injection, the minimum required aperture is 30 mm, so the sensors need to be retracted. Each station is therefore divided in two halves, each halve covering slightly more than 180° to obtain full

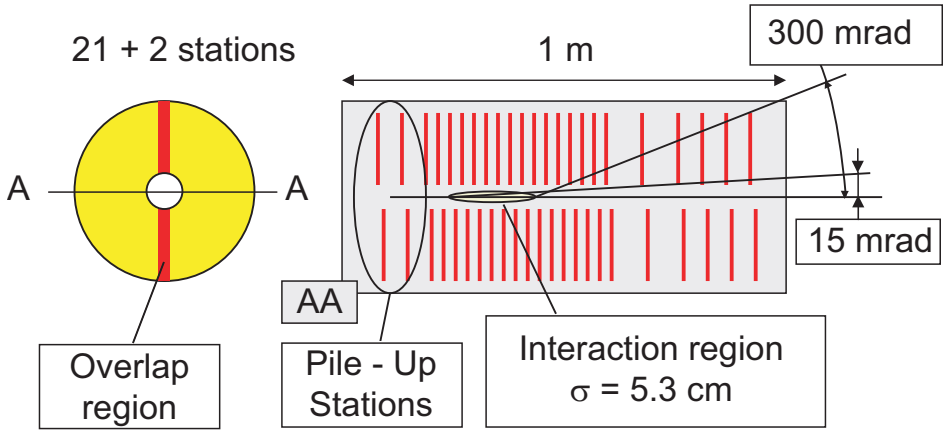


Figure 3.18: Schematic sensor arrangement in the vertex detector.

ϕ-coverage and to accommodate for alignment.The complete vertex detector can be centered around the beams with 2 translation frames before the detector halves are moved into the positions required for B-physics data taking. In total there are 21 stations for the vertex detector. Two extra stations are mounted in the backward direction outside the acceptance, that constitute the pile-up detector [47].

A box made of thin aluminum foil shields the silicon sensors from RF radiation from the beams. It also acts as a wakefield guide [49] and avoids pollution of the beam vacuum due to outgasing of cables and electronics. The complex shape of this foil is shown in Fig. 3.19. It accommodates for overlap between the detectors and minimizes the average amount of material that the particles traverse.

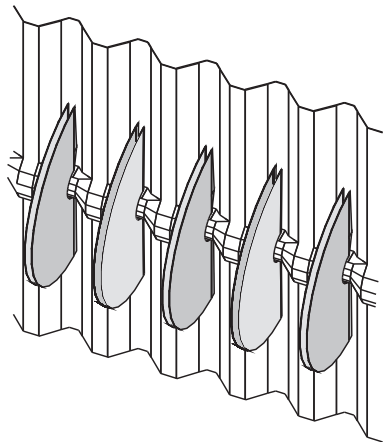


Figure 3.19: Part of the foil with five sensor pairs. The complex shape accommodates for overlap between the detectors and minimizes the average amount of material that the particles traverse.

The foil has a nominal thickness of 250 μm and is produced with super plastic hot gas formation, a technique that allows for extreme deformations in thin foils. It can withstand a pressure difference up to 15 mbar without plastic deformation, which implies that the pressure on both sides needs to be monitored and controlled at all times. The simplest and

most stable mode of operation is by maintaining a secondary vacuum in the volume that contains the silicon sensors [50]. A cross section of one of the detector housings is shown in Fig. 3.20. The foil and the bellows needed for retraction of the sensors are indicated.

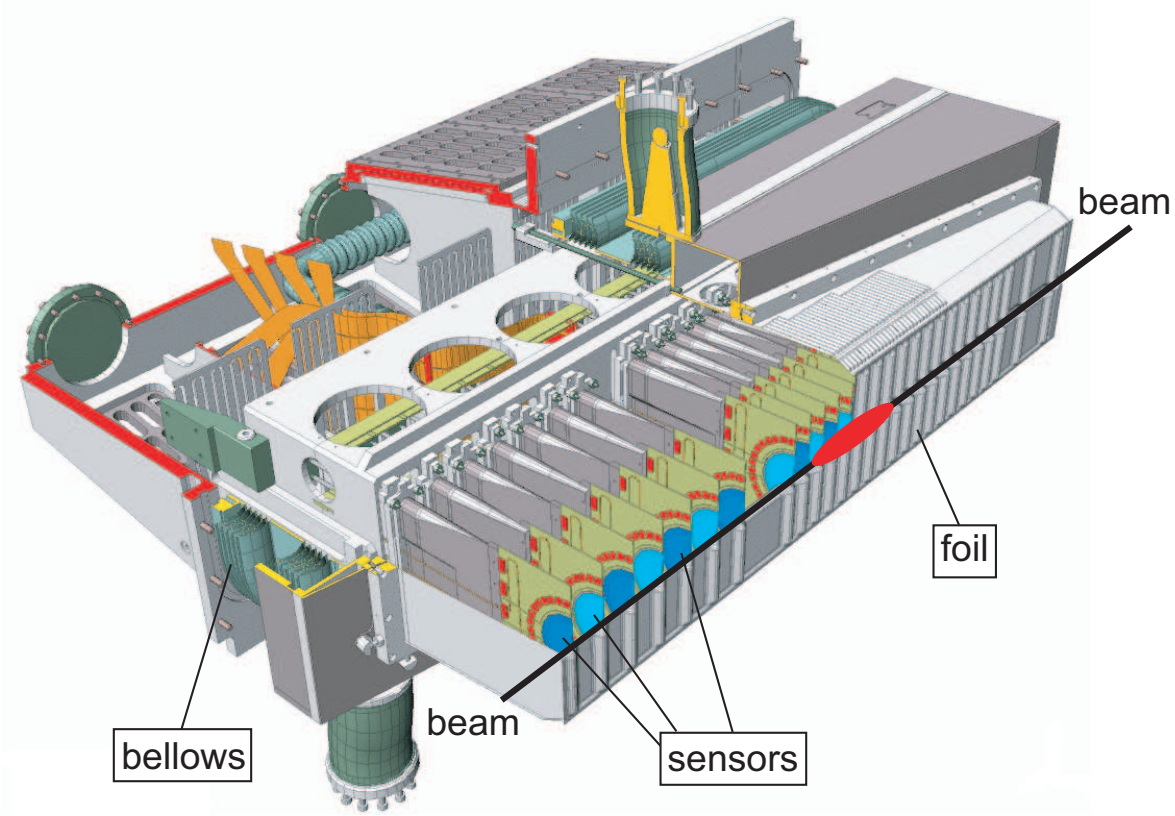


Figure 3.20: Cross section of the silicon strip detector housing.

Maintenance of the vertex detector is a delicate operation. The system needs to be brought to atmospheric pressure before the volume with the silicon stations is accessible. During this transient state the pressure difference over the foil should not exceed the specified value. This is accomplished by slow and simultaneous venting of the secondary and beam vacuum with ultra pure inert gas, which preserves the secondary electron emission properties of the NEG coating (see section 3.1).

If one has to break the seal between beam and detector vacuum, then the beam vacuum system needs to be baked out before LHC operation can be resumed. Before this bake out the silicon stations need to be removed, since the tank is heated to 120°C, the exit window to 150°C and the foil to 160°C. A second venting and evacuation cycle is required to install them again. A detailed description of the PLC control logic for all pumps, valves and sensors involved in the transient vacuum procedures of the vertex detector is given in appendix 7.

Note that the bake out of the vacuum system does not only affect the vertex detector. The beam pipe through the rest of the experiment needs to be baked out at temperatures between 200 and 250°C. The effects of the bake out on the glued kapton seal of RICH1, which is mounted directly on the exit window of the vertex detector, are still under investigation. The TT as well as the inner and outer tracker stations can be split in two halves. Removable

heating jackets need to be installed for the bake out of this part of the beam pipe. The temperature of the inner wall of RICH2 should stay below 70°C. Tests showed that this requirement is met with removable heating jackets, which are installed in the 45 mm gap between this sub-detector and the beam pipe. The SPD and PS calorimeters have the same aperture as RICH2, but they can be split in two halves during bake out. The rest of the calorimeters and the muon chambers have a fixed bake out system.

3.3.2 Silicon sensors

The 300 μm thick silicon micro-strip sensors are shown in Fig. 3.21. Each station has two sets of sensors with R-strips and ϕ-strips mounted back to back. The inner radius of the active area of the sensor is 8.0 mm and the outer radius amounts to 42.2 mm. The pitch varies from 40 μm (32.5 μm) in the inner to 101.6 μm (96.6 μm) in the outer region for the R (ϕ)-sensor. The combination of an $R\phi$-geometry with the smallest pitch at small radii leads to the highest XY-resolution for tracks closest to the interaction region. The ϕ-strips have a dog-leg shape due to the stereo angles of -20° in the inner region and +10° in the outer region. The stations are alternately installed with the R- and ϕ-strips toward the interaction point, which helps to resolve ambiguities when R and ϕ information of different tracks is combined. A typical event in *LHC*b has about 70 tracks.

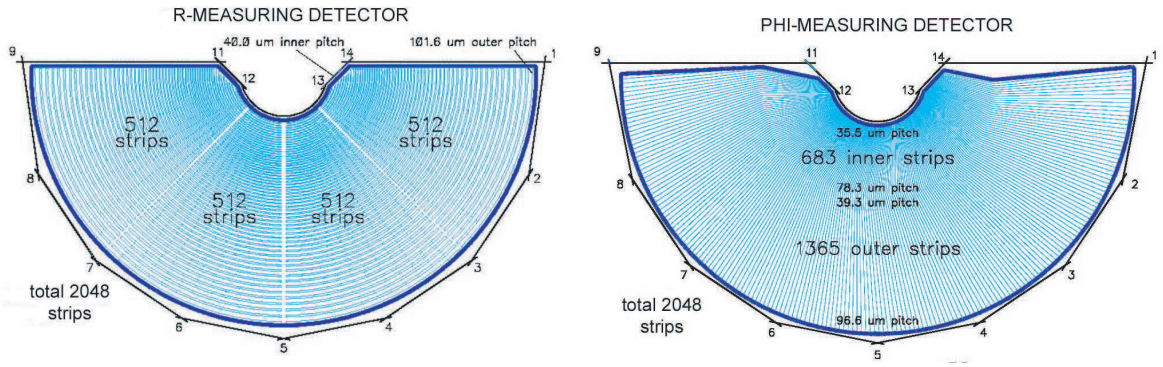


Figure 3.21: R *and* ϕ *sensor.*

As sensor material n-on-n oxygenated silicon was selected. A simplified sensor cross section is shown in Fig. 3.22. The electrons of the n-doped bulk fill up the holes in the p-doped backplane, which results in a small depleted region on the back plane side when no bias voltage is applied. Applying a sufficiently large reverse bias voltage across this junction depletes the bulk completely by accumulating all available electrons on the n^+-doped strips and the holes on the backplane. A disadvantage of n-on-n sensors is the technical complexity. Special measures, like p-stops, are needed to prevent electrical shorts of the n^+-strips.

Charged particles that traverse the detector generate electron-hole pairs in the bulk material. The electrons are collected on the strips and form the signals, which are routed via a double-metal layer to the read-out system. It is an important advantage of the n-on-n configuration that the signal is mainly formed by electrons, because they move faster than holes and will therefore provide shorter pulses.

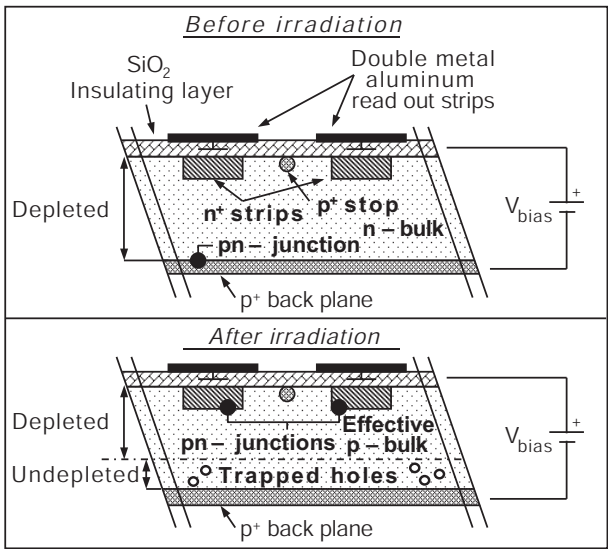


Figure 3.22: Simplified sensor cross section, showing the n-on-n silicon sensor in its original state and after irradiation.

Another advantage of the n-on-n configuration is related to radiation damage. Lattice displacements caused by non-ionizing energy loss interactions result in the creation of additional acceptor states. The n-doped bulk is slowly converted into p-type, a process that is known as type inversion [51]. When this happens the junction will form between the n^+-doped strips and the effective p-bulk. The bulk gradually becomes even more p-type and the increase of positive space charge needs to be compensated by raising the bias voltage. This solution is limited by the break down voltage of the detector, which means radiation damage will eventually lead to a reduced depletion depth. The undepleted layer acts as an insulator in heavily irradiated silicon, thus trapping the holes that should be collected on the back plane [52]. In the case of an n-on-n configuration the insulating layer reduces the signal height but not the resolution, since the undepleted region is not on the strip side. Therefore, hit identification stays close to 100% efficient even when radiation damage reduces the depletion depth to 60%, as shown on the left hand side of Fig. 3.23.

The expected maximum dose is $1.3 \cdot 10^{14} \mathrm{n_{eq}cm^{-2}}$per year and varies slightly from station to station, as shown in the inset on the right hand side of Fig. 3.23. The irradiated detectors have to be capable to withstand the high bias voltages needed for depletion. The formation of an undepleted layer will result in an even larger gradient over the rest of the bulk. Furthermore, the r dependence of the radiation intensity, also shown in Fig. 3.23, poses an additional challenge: the non-uniform radiation intensity can lead to a situation where part of the sensor will have the pn-junction on the backplane side, while the other part has it on the strip side. Beam tests showed that newly developed n-on-n sensors can operate correctly under these conditions and that they will be able to survive about 4 years of *LHC*b operation.

A radiation damage related effect is annealing: the lattice displacements slowly disappear over time, which partially reverts the type inversion. However, above a certain temperature the opposite effect, *i.e.* accelerated radiation damage known as reverse annealing, occurs as well. Since the lattice displacements are acting as generation-recombination centers for electron-hole pairs, they do not only cause type inversion, but also result in an increased leakage current. The cycle: reverse annealing → accelerated radiation damage → increased

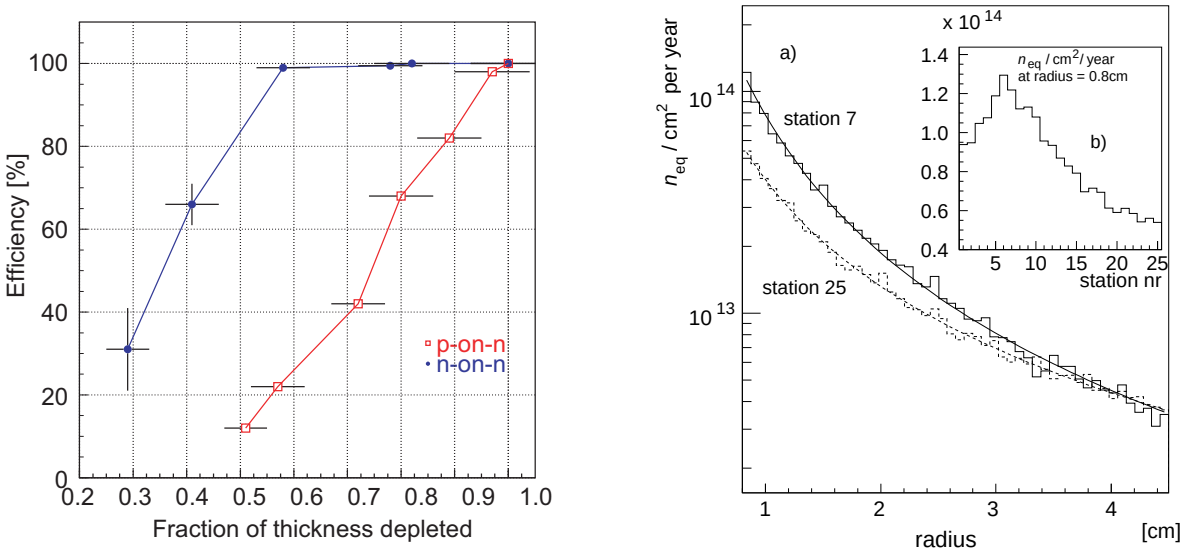


Figure 3.23: On the left: Hit identification efficiency as a function of depletion depth for p-on-n and n-on-n sensors. On the right: expected radiation dose of the vertex detector stations for 1 year of operation.

leakage current → increased power consumption → increased temperature → more reverse annealing, is called 'thermal runaway'. To avoid reverse annealing, the temperature of the sensors should be kept below $-5°$C.

The cooling of the sensors is provided by evaporation of CO_2 in a closed-loop cooling system. This system is mounted as much as possible outside the acceptance and high conductivity carbon (TPG − Thermal Pyrolytic Graphite) is used as a support for sensors and hybrid to conduct the heat with a minimal amount of material. On the left hand side of Fig. 3.24 the part of the cooling system inside the detector acceptance is shown.

The chip temperature will be considerably higher then the rest of the system, since the chip is a heat source and the conductivity of TPG is high in one direction, but very low in the others. This anisotropy also leads to critical connections between the cooling tubes and the TPG, which should provide sufficient conductivity with a minimal amount of material. These connections are still under investigation.

On the right hand side of Fig. 3.24 an overview of the cooling components inside the vertex detector vacuum is given. The CO_2 is transported as liquid at room temperature under high pressure. The cooling cycles can be visualized with a Mollier diagram[8] as shown in Fig. 3.25. As soon as the pipe with liquid CO_2 enters the vacuum tank it goes through a heat exchanger (A), using the return gas for pre-cooling to a temperature of about $-30°$C. After the heat exchanger, the liquid enters a manifold where it is equally distributed over the cooling tubes (*i.e.* the flow resistance of the tubes is much larger than the resistance of the manifold). When the liquid enters the tubes it passes restrictions and is expanded adiabatically (B) to obtain the right combination of pressure and temperature. During the expansion the CO_2 is only partially evaporated, making sure that the liquid/gas mixture cools

[8] A Mollier diagram shows the relation between pressure, temperature and energy content of a substance. Transient thermodynamic processes (*e.g.* heat exchange and expansion) can be analyzed with these diagrams.

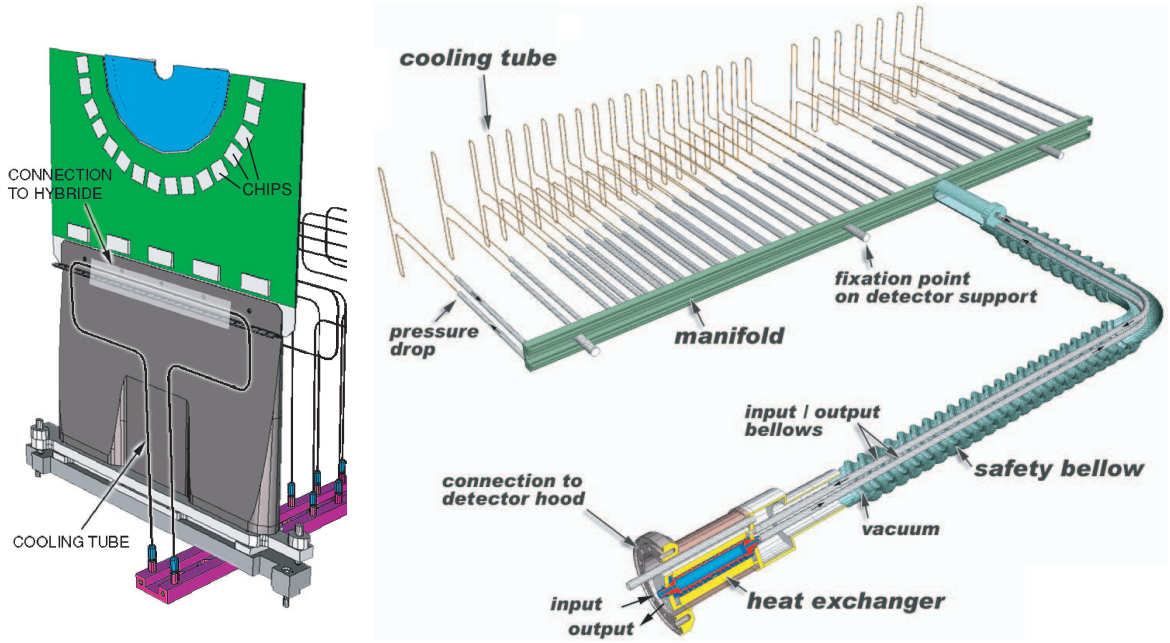


Figure 3.24: On the left: silicon detector module. High conductivity carbon supports the hybrid with the sensor. On the right: schematic overview of a part of the closed-loop CO_2 cooling system.

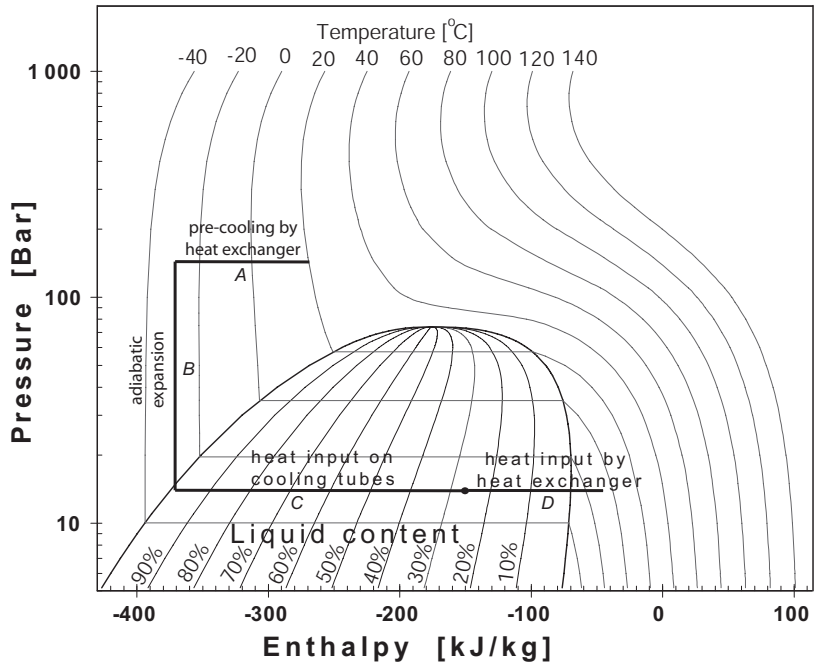


Figure 3.25: Mollier diagram of the CO_2 cooling trajectory. Pre-cooling of the pressurized liquid is indicated with A. Then, after the adiabatic expansion indicated with B, the vapor enters the tubes and provides cooling capacity by means of evaporation (C). The remaining vapor pre-cools the incoming liquid (D).

the modules without a temperature gradient (*C*). The return gas is fed to the heat exchanger where it pre-cools the incoming liquid (*D*). After this, it is recycled in an external cooling system outside the vacuum vessel.

3.3.3 Read-out chip

The Beetle is a 128 channel radiation hard read-out chip with an analog pipeline that is designed to operate at 40 MHz. The chip is based on 0.25 μm CMOS integrated circuit technology and implements the RD-20 front-end architecture [53]. The chip can withstand high radiation doses and almost no deteriorating effects were observed up to 300 kGy (corresponding to about 15 years of *LHC*b operation) [54]. The Beetle has originally been designed for the *LHC*b vertex detector, but will also be used by the pile-up detector, the trigger tracker and the inner tracker of *LHC*b.

The response of the Beetle to electrical signals in the strips of the silicon sensor (caused by charge depositions of minimum ionizing particles in the detector) is mainly determined by the shape of the input signal and the transfer function of the front-end circuit of the chip. As explained in the previous section, the input signal is formed by collection of electrons. These pulses are, with sufficient bias voltage over the detector, much shorter than the output pulse generated by the Beetle and can be considered as being delta functions.

A functional diagram of the Beetle is shown in Fig. 3.26. It contains the front-end circuit, a pipeline cell and the read-out amplifier of one channel. The dummy channel is shown as well, which is the same as the other channels, but without a pre-amplifier. With this channel the output amplifier corrects for common-mode noise that originates in the chip. The settings of the Beetle are controlled via an I^2C interface[9]. Inter-IC or I^2C is a control, diagnostic and power management bus, which allows programming of the chip registers at speeds up to 3.4 Mbit $\cdot$ s^{-1}.

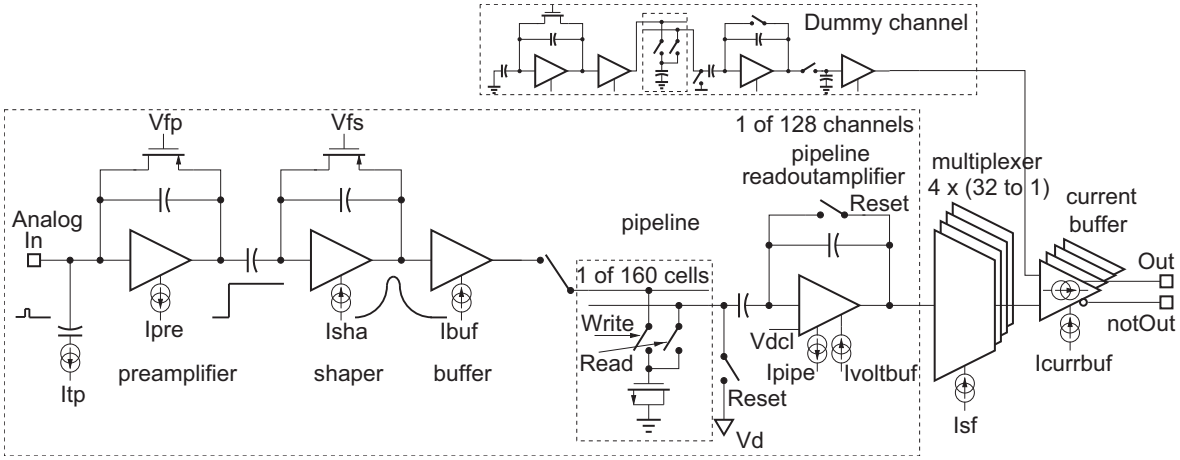


Figure 3.26: Functional diagram of the Beetle. The front-end circuit, a pipeline cell and the read-out amplifier of one channel are shown. The dummy channel, in the upper right corner, provides common mode correction. It is the same as the other channels, but does not contain a pre-amplifier.

[9] I2C-bus specification, http://www.semiconductors.philips.com/acrobat/literature/9398/39340011.pdf

The pre-amplifier integrates the input signal, which in case of a delta function results in a step function. This step function gives a better measure of the energy deposition in the detector and reduces the sensitivity to noise picked up by the detector strips. The noise of the output signal of the Beetle is dominated by the pre-amplifier circuit, since its amplification is much higher than that of the subsequent shaper circuit[10]. This noise can be minimized by increasing the pre-amplifier current (I_{pre}) up to the limit given by power consumption considerations[11]. Note that the steepness of the rising edge of the step function created by the pre-amplifier depends on the detector capacitance and the impedance of the pre-amplifier circuit. This impedance is also affected by changes in the pre-amplifier current. Increasing this current does therefore not only reduces the noise, but also reduces the rise time of the output signal of the pre-amplifier.

The signal from the pre-amplifier needs to be shaped to obtain the best signal to noise ratio on one hand and to be able to separate pulses which are 25 *ns* apart on the other hand. The shaper consists of a low pass filter (integrator) and a high pass filter (differentiator), the time constants of these filters can be tuned with the feedback resistance of the shaper (V_{fs}) and the shaper current (I_{sha}).

After the signal is processed, it is sampled at 40 MHz and the results are stored in a 187 cells deep pipeline. The corresponding maximum trigger latency is 4μs. A buffer, implemented as a simple source follower, provides the low impedance to drive this pipeline (or the comparator in case of digital read out by the pile-up system). If an event needs to be read out, the pipeline column number is stored in the so-called derandomizing buffer and the stored signals are retrieved from the pipeline. The signals are multiplexed in 4 groups of 32 channels and sent in parallel via a full differential current output to the off-detector electronics. The transfer of all signals takes $(32+4\,header\,bits)\times 25\,ns = 0.9\,\mu$s, allowing for continuous read-out at 1 MHz as required by the level-1 trigger.

The 16 cells deep derandomizing buffer allows the Beetle to cope with fluctuations in the trigger rate. The addresses of the pipeline columns that need to be read out will stay in this buffer until they are retrieved from the pipeline. The write pointer skips the column numbers stored in this buffer and thereby prevents that the data are overwritten. As a result, the Beetle is able to temporarily accept a trigger rate larger than 40 MHz, as long as the derandomizing buffer has slots available.

Note that it is also possible to multiplex all 128 channels and send them in 1 group to the off-detector electronics. This mode was used in the beam test described in chapter 4, because only a limited number of ADC channels was available for read-out. However, the level-1 trigger specifications are not met in this mode, since the transfer of all 128 signals takes more than 3.2 μs. A detailed description of the Beetle chip can be found in the Beetle reference manual[12].

An early version of the chip was extensively tested at the SPS, the 120 GeV pion beam

[10] Note that the pre-amplifier is a transconductance amplifier and the shaper circuit is a voltage amplifier, so their properties can not be compared directly. Nevertheless, the amplification of the pre-amplifier is sufficient to avoid significant noise contributions from the shaper circuit to the output signal.

[11] Actually, the noise is inversely proportional to the transconductance (g_m) of the input transistor. This transconductance increases linearly with the current that flows through this transistor, which is determined by the pre-amplifier current.

[12] http://wwwasic.kip.uni-heidelberg.de/lhcb/Documentation.html.

facility at CERN. A prototype hybrid was equipped with 16 Beetle1.1 chips and a 300 μm thick silicon sensor was used as detector [55]. The most probable energy deposition of a minimum ionizing particle in such a detector is about 80 keV [51], which means that on average 22,000 electron-hole pairs are created. Combined with an ENC relation measured for this specific chip version of $871e^- \pm 41.5e^- \cdot \mathrm{pF}^{-1}$ and an estimated detector capacitance between 10 and 20 pF (including routing lines, pitch adapter and bonds) [56], this results in expected signal/noise ratios between 13.5 and 17.5. Note that the chip settings during this ENC measurement were not at all optimized, so these numbers provide an indication only. The beam-test is described in detail in chapter 4.

The Beetle1.3 contains various bug fixes that should further improve the performance. A fully equipped hybrid with 16 Beetle1.3 chips is tested in the SPS beam in the first half of 2004. Two more iterations of this chip are tested, mainly to improve the threshold accuracy of the digital output for the pile-up detector and to remove some oscillatory behavior. Based on the results with these new versions it will be decided whether the 1.3, 1.4 or 1.5 will be installed in the vertex detector.

Chapter 4

Characteristics of 16 Beetle1.1 chips on a prototype vertex detector hybrid

Here, the results of beam-test measurements performed with a prototype vertex detector hybrid equipped with 16 Beetle1.1 chips in the 120 *GeV* pion and muon beam at the X7 facility[1] at CERN are presented. A picture of the experimental setup is shown in Fig. 4.1.

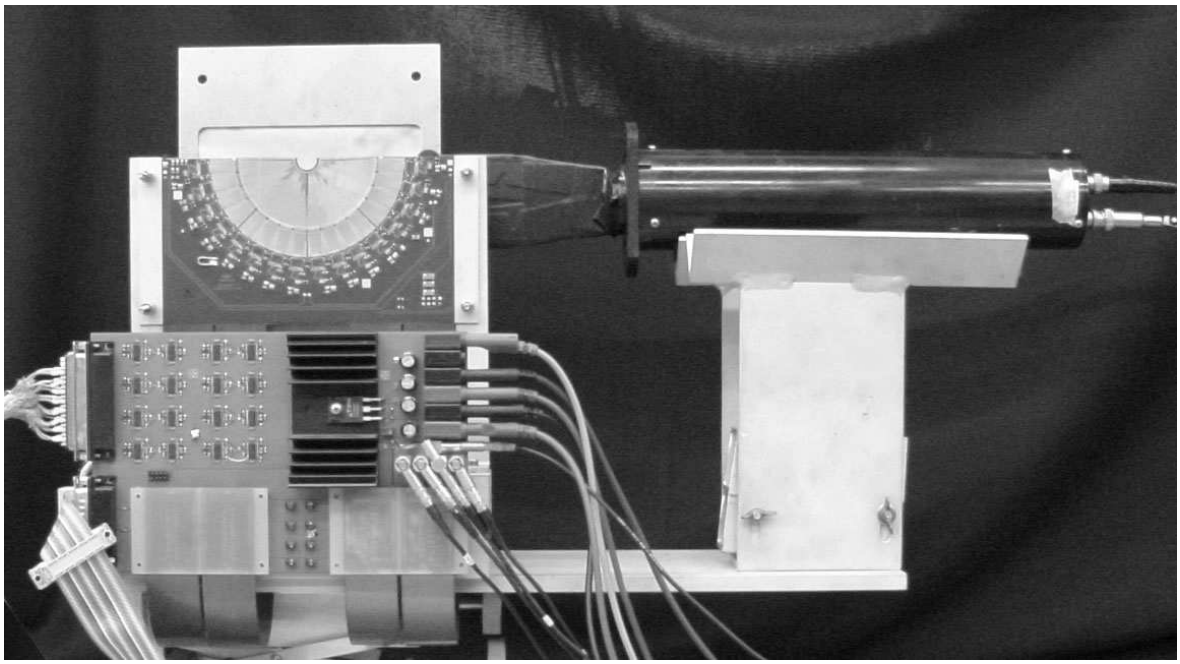


Figure 4.1: Photograph of the experimental setup. One of the scintillators blocked the view and was removed for the photo.

The hybrid is connected to a 300 μm thick p-on-n silicon detector. The purposes of the experiment were to study the system response to minimum ionizing particles (MIPs) and to use 16 chips on a single hybrid. The front-end settings of the chip were tuned to meet the *LHC*b vertex detector requirements. Channel-dependent effects were studied with data from

[1] http://lhcb-tb.web.cern.ch/lhcb-tb/html/x7.htm

high statistics measurements. High trigger rate and single time sample measurements were analyzed to investigate specific aspects under *LHC*b conditions.

The Beetle has several parameters that can be used to tune pulse shape characteristics like rise time, signal/noise ratio and spill-over. The most important parameters are the pre-amplifier current (I_{pre}), the shaper current (I_{sha}) and the feedback resistance of the shaper (controlled by V_{fs}) [2]. The pre-amplifier was operated at its maximum possible current of 350 μA, as explained in section 4.1. Based on results from previous experiments [57] a shaper current of 32 μA was selected to create a pulse shape with at maximum 15% undershoot. This is a compromise between the height and the tail of the pulse. An overview of the data is presented in Table 4.1.

Table 4.1: All data sets as measured in the experiment. I_{sha} = 32 μA, I_{pre} = 350 μA.

	Type	V_{fs} [mV]	Events [M]
1	Pulse shape scan, chips 8, 9, 10, 13, 14, 15	500-800	1.4
2	High statistics runs, chips 8, 9, 10, 13, 14, 15	700	1.0
3	Pulse shape scan, all chips	500-800	1.6
4	Single time sample runs	700	1.0
5	High trigger rate runs	700	1.0
	Total		6.0

Data set 1 was used to tune the feedback resistance of the shaper (V_{fs}) and to find an optimized pulse shape for the vertex detector. Data set 2 represents a high statistics measurement and is used to investigate channel dependencies and determine systematic errors. The stability and reproducibility of the measurements can be studied for all chips with data set 3. Data sets 4 and 5 correspond to the single time sample and high trigger rate measurements respectively. Both are used to investigate specific aspects of *LHC*b operation and to cross-check the results.

This chapter has the following structure: section 1 gives a description of the setup used in the beam-test. In section 2 noise studies are presented. Section 3 discusses the pulse shape parametrization and section 4 presents the pulse shape analysis including all cross-checks and results.

4.1 Experimental setup

A prototype vertex detector hybrid has been designed at NIKHEF and was constructed by Eltech[3] to test 16 Beetle1.1 chips in the X7 facility at CERN. This facility provides a tertiary particle beam containing pions and/or muons with a typical energy of 120 GeV. The parent beam, extracted from the SPS ring, delivers about 10^7 particles (pions and electrons) per spill of 5 seconds to the secondary targets of the west area test beam complex. The spot size of this beam is a few millimeters RMS in each transverse direction. The copper target converts the beam to pions and muons. Optionally, the pions in the beam can be stopped with

[2] Beetle reference manual, http://wwwasic.kip.uni-heidelberg.de/lhcb/Documentation.html.
[3] http://www.eltech.com

a concrete block, in which case only muons will survive. The results include measurements with both opened and closed beam block.

For the experiment the system was extended with two scintillators and an XY station read out by HELIX chips [58], as shown in Fig. 4.2. The Beetle off-detector electronics was close to the beam, the rest of the data acquisition system was about 8 meters away.

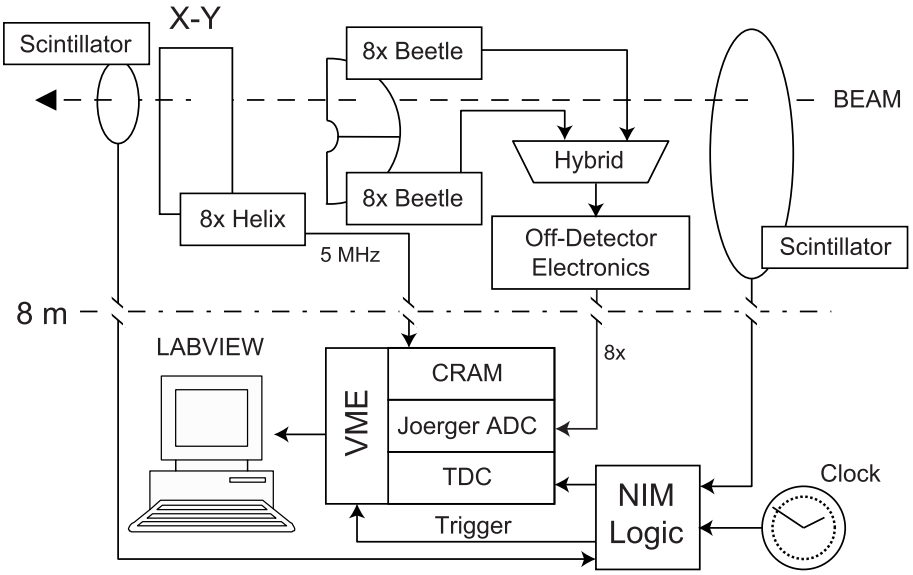


Figure 4.2: Overview of the setup used in the beam-test.

A 300 μm thick PR02-R p-on-n detector was used as sensor. It consists of 2,048 strips divided in 16 sectors as shown in Fig. 4.3, each of which is read-out by one Beetle chip. The inner and outer radius of the detector are 8.0 and 42.2 mm, respectively. The strip width is about 40% of the pitch, which increases from 32.5 μm in the inner regions to 92 μm in the outer regions. The detector was biased at 150 V which is 110 V above the depletion voltage. We refer to this detector as 'the Beetle detector'.

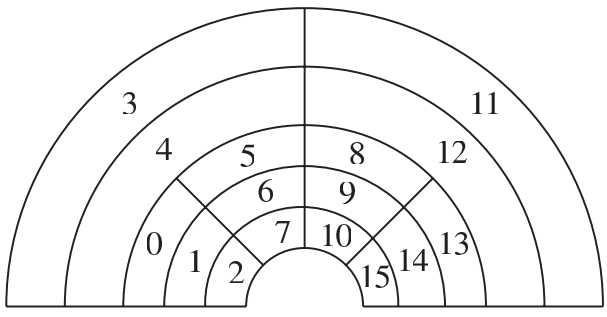


Figure 4.3: Layout and sector numbering of the Beetle detector.

A two layer pitch adapter is used as a transition between the detector pitch (2 rows of bond pads, effective pitch = 61.5 μm) and the chip pitch (4 rows of bond pads, effective pitch = 40.24 μm). The pitch adapter consists of a 200 μm thick epoxy substrate on top

of which a 50 μm thick kapton layer is glued. The pitch adapters were glued on identically shaped 200 μm thick aluminum plates to flatten them, before they were mounted on the hybrid. By accident the aluminum plates are left electrically floating, which introduces additional noise (see section 4.2.1).

The 16 Beetle1.1 chips were bonded without prior testing. They are all fully operational, which indicates a high yield of the associated production run in deep sub-micron technology. They are read-out in single port mode, which means when a trigger arrives the signals from 128 channels are serialized to reduce the number of ADC channels needed. Each read-out starts with an 8 bit header specifying the pipeline column number. A term used through out this chapter, related to the chip read-out, is 'time sample': a 'time sample' contains the signals of all 128 chip channels that belong to the same trigger. Two consecutive 'time samples' are 25 ns apart (corresponding to a 40 MHz clock cycle).

The pre-amplifier current, the shaper current and the feedback resistance of the shaper are programmed with an I^2C interface[4] via 10 bits on-chip Digital to Analog Converters (DAC). The pre-amplifier is operated at its maximum possible current of 350 μA[5] .

The hybrid consists of a four-layer printed circuit board with kapton insulation layers of 75 μm thickness and a copper thickness of 17 μm. Connections between the hybrid and the off-detector electronics (ODE) board are made via flexible cables called tails, one of which was fabricated from solid copper and the other one from copper with a grid. On the ODE board, about 30 cm away from the Beetle chips, the line-drivers amplify the signals by a factor 36. A group of 8 signals can be amplified simultaneously; jumpers are used for selection.

The silicon sensors need to be operated in a light-tight enclosure. Cooling is enforced by circulating air in the box with a ventilator to obtain an operational temperature of about 35°C. The signals are sent over 8 meter long cables to the 8 channel, 12 bit ADC VME-module (Joerger VTR812-40). Two broken ADC-inputs reduced the number of available ADC channels to 6.

An XY tracking station with a surface area of 6×12 cm^2 provides an additional space-point for each track. This tracking station consists of two detector planes mounted back to back with a strip pitch of 20 μm and a read-out pitch of 120 μm. Read-out is done by HELIX chips connected to a CAEN V550 10 bit VME-ADC (CRAM) operating at 5 MHz.

Triggers are constructed from a coincidence between the signals from two scintillators. A TDC measures the time difference between the trigger and the rising edge of the next Beetle clock pulse. Since the beam is continuous in the spill, the TDC information can be used to reconstruct the transient response of the Beetle chip. For the majority of the data, a single trigger issued the read-out of 8 consecutive time samples, spanning a total period of 200 ns. In this way the complete pulse shape is reconstructed from the data, without the need to change the latency. A data acquisition program written in LABVIEW, controls a non-graphical C-based program capable of collecting up to 500 events per spill of 5 seconds.

[4] I2C-bus specification, http://www.semiconductors.philips.com/acrobat/literature/9398/39340011.pdf

[5] This pre-amplifier current limit is due to saturation of the DAC circuitry in the Beetle1.1. The corresponding DAC settings are: DAC_{pre} = 600 (for a pre-amplifier current of 350 μA) and DAC_{sha} = 50 (for a shaper current of 32 μA). The Beetle1.2 shows no saturation up to the maximum design value of 2 mA for the current.

4.2 Noise studies

In order to correctly interpret the data it is important to understand the noise characteristics of the system. Measurements of the noise give a first indication about the quality and uniformity of the setup and the chip response. Noise originates in the detector, the Beetle chip, the line-drivers and in the Joerger ADC that is used in the data acquisition. In addition, noise can be picked up by the hybrid, the pitch-adapter and by the cables that connect the Beetle to the Joerger ADC. In the plots of this section, measurements with Beetle settings I_{sha} = 32 μA, I_{pre} = 350 μA and V_{fs} = 700 mV are shown.

4.2.1 Total noise

The distribution of the pedestal-subtracted signal of a time sample that precedes the samples that contain the pulse, is presented in Fig. 4.4. The noise is highly symmetric and Gaussian distributed with a normalized χ^2 of $\approx$ 1.09. The distributions are qualitatively the same for all Beetle chips at all settings. The standard deviation of the Gaussian is called 'total noise', which is determined for each channel and each time sample separately.

For every trigger, the ADC measures 24 samples before the data from the Beetle arrive. These samples can be used to determine the so-called 'system noise', *i.e.* noise generated in the Joerger ADC or by the line-drivers. This noise depends on the Joerger input used, and varies from file to file. Typically it is around 6 ADC counts, with the exception of Joerger input 7 which has been used to read-out chip 7 and 15, where the noise varies between 7 and 14 ADC counts due to ground loop currents. No attempt was made to subtract this noise.

Fig. 4.5 shows the standard deviation of the distribution of the pedestal-subtracted signal as a function of channel number for chip 1. The distributions of the other chips look similar. Some features are apparent:

- The first 4 channels of the chip have a high total noise. Histograms with the raw data of these channels show two peaks, which are caused by a pedestal that is correlated with the sign of the last header bit. More detailed studies showed that the signal of the first channel contains an oscillation that fades out when the sample time is delayed. This ringing of the external amplifiers is presumably caused by insufficient phase margin or gain bandwidth [59]. The external amplifiers used, are operated at their limit because of the small output signal of the Beetle1.1. An independent measurement showed that reflections on the analog read-out cables disturb the signal of the first four channels. A tuned system (impedance, amplifiers and signal termination) in combination with the four times larger amplification factor of future chip versions will eliminate this problem. From here on data of the channels 0, 1, 2 and 3 of all chips are rejected in the analysis.
- Some chips have channels with a total noise around 12 ADC counts and/or with much higher noise than the average. The channels with low noise correspond to unbonded channels, the channels with high noise probably suffer from cross-talk with the neighboring routing line or bond. In section 4.2.2 more detailed information about these anomalous channels can be found.
- The total noise as a function of channel number has a double band structure: in a group of 4 consecutive channels, the first two channels have on average about 10%

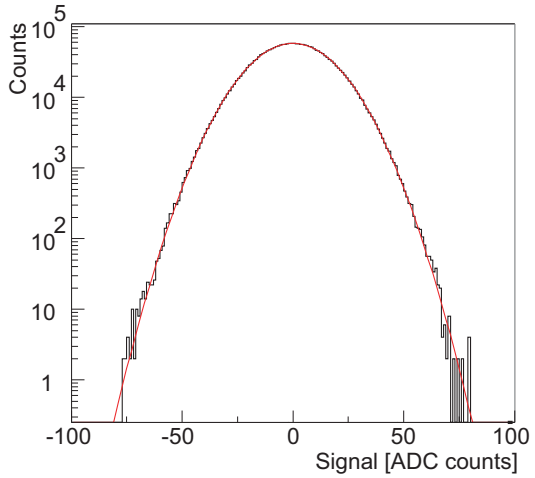


Figure 4.4: Pedestal-subtracted signals for 10k events in Beetle chip 0. The distribution is fitted with a Gaussian.

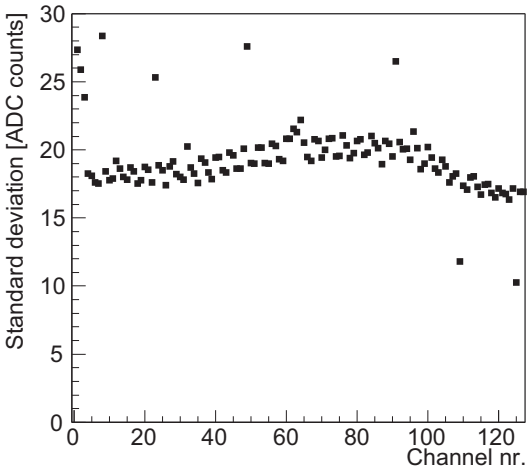


Figure 4.5: Total noise for 10k events in chip 1 versus channel number.

more noise than the last two. This structure matches the layout of the pitch adapter, where the first two channels of each group of 4 are routed over the lower layer and the last two over the upper layer. The higher noise in the lower layer is due to the floating aluminum plate on which the pitch adapter was glued. This issue was further investigated via two independent measurements. First, a high frequency oscillator generated a 5 MHz RF signal at the open end of a coaxial-line positioned at the center of the detector. The analysis results of this measurement showed a similar double band structure as shown in Fig. 4.5. Second, two bonds were swapped on the chip side of the pitch adapter and two bonds of other channels on the silicon side. In the former case the additional noise moved to a different channel, while in the latter nothing changed. This indicates that the RF pick-up as well as the noise leading to this double band structure originate in the pitch adapter.

- The noise of chips connected to sensor regions with long strips (chip 3, 4, 11 and 12) is up to 10% higher than the noise of chips connected to the strips in the inner region, which are shorter and therefore have less input capacitance [56].

In order to compare the absolute noise values for different chips, a gain factor was determined based on the difference in ADC value between high and low header bits. This difference is independent of chip settings and was in our measurements about 80 times larger than the width of the noise distribution. The gain factors are typically between 0.85 and 1.15. From here on, the noise that is shown is gain-corrected.

4.2.2 Noise anomalies

Channels with an exceptional noise behavior were excluded from the analysis. Channel numbers 0 to 3 are excluded for all chips due to ringing in the external amplifiers and reflections on the analog output cables as discussed in section 4.2.1. Channel numbers 0, 32, 64 and 96 are excluded for all chips because they pick up a few percent extra noise due to a multiplexer

feature (which is solved in the Beetle1.2)[6]. For 34 channels the noise level was around 12 ADC counts (34 channels). An RF generating setup was used to study the response of these channels. The test showed that these channels do not respond to the RF signal. This supports the conclusion that these channels suffer from broken bonds instead of problems with chip internals or read-out electronics. Another 41 channels have extra noise due to shorts with the neighboring channels in either the bonds, the silicon sensor or the pitch adapter. It was verified that the noise of these channels returns to a nominal value when the bond of the neighboring channel is removed. On average we excluded 11.7 channels per chip (9.1%). Some of the chips have no additional anomalous channels, while others have as much as 16 extra channels with problems. The data of the channels listed in Table 4.2 were excluded from the analysis due to anomalous noise effects.

Table 4.2: Overview of all excluded channels.

Chip number	Excluded channels
All	0, 1, 2, 3, 32, 64, 96
0	56, 57
1	8, 23, 49, 62, 63, 91, 109, 125
2	12, 24, 56, 125
3	11, 20, 24, 25, 34, 36, 40, 48, 71, 72, 101
4	4, 8, 41, 42, 43, 57, 68, 108, 109, 112
5	5, 120, 121
6	29, 74, 84
7	80
8	
9	17, 25, 73
10	84, 104, 106
11	5, 7, 9, 12, 16, 18, 19, 24, 28, 30, 57, 77, 81, 109, 112, 122
12	48, 49, 92, 125
13	25, 41, 72, 74, 88, 112
14	
15	92

4.2.3 Common-mode noise correction

Subtraction of common-mode noise is required to correct for low-frequency noise that is picked up by the read-out electronics or the silicon detector. Four different strategies were investigated. Common to all four strategies is the exclusion of channels that contain a hit. In this case a channel is considered to contain a hit if its ADC value is larger than 5 times the total noise of that channel and time sample. The four different strategies are described below.

[6]http://wwwasic.kip.uni-heidelberg.de/lhcb/Meetings/Meeting220103/Beetle_Lab_Measurements.pdf

1. The average pedestal subtracted signal of the channels of one chip is subtracted from the data.
2. The pedestal subtracted signal is fitted with a first-order polynomial as a function of chip channel number. This corrects for a common-mode picked up between the chip and the ADC that slowly varies in time ($\sim$ kHz), since the channels are read-out sequentially.
3. The pedestal subtracted signal is fitted with a first-order polynomial as a function of strip length. This corrects for a linear dependence of common-mode pickup on strip capacitance.
4. Individual weight factors are determined for each channel by constructing a spectrum with on the x-axis the common-mode corrected noise in a channel and on the y-axis the common-mode noise as calculated with method 1. A straight line is fitted to the average values to determine the correlation. The correction factor is one plus the slope of this line. If this slope equals zero, the channel is weighted by a factor 1 and it picks up just the average amount. If the slope equals -1, the channel picks up no common-mode noise at all.

The common-mode noise for a chip is Gaussian distributed, with a σ of around 10 ADC counts. Correction of the data for common-mode noise reduces the noise per channel with about 13%. This effect slightly differs from file to file and from chip to chip. The extra reductions that can be obtained by applying the strategies 2, 3 and 4 amount to about 1%, 0.6%, and 2%, respectively. For strategies 2 and 3 this is to be expected from the fact that a second fit parameter was introduced, hence the improvement should be in the order of $N_{d.o.f.}^{-1} \approx 1\%$ (where $N_{d.o.f.}$ represents the number of degrees of freedom). Table 4.3 summarizes the results, obtained by analyzing 40k events.

As mentioned before, the straight-forward common-mode subtraction performed in strategy 1 already yields a close to optimal result; more involved strategies like strategy 4 give negligible improvement. However, for historical reasons[7] strategy 2 is applied in the rest of the analysis: instead of a constant factor, a straight line is fitted through the pedestal-subtracted ADC data as function of the chip channel number.

In the rest of this section the channel to channel variations in the noise are discussed in detail. Figure 4.6 shows the total noise and the noise corrected for common-mode with strategy 2. A clear channel dependence is visible. As in the pedestal-subtracted signal, the first two channels in a group of 4 are noisier than the last two. Obviously, fitting a constant or straight line common-mode term does not remove this feature but blurs it a bit. Furthermore, after common-mode correction the noise of channels with channel numbers above 120 increases. These features are observed in all settings studied. During an independent measurement the Beetle was programmed to send only the data of the first 32 channels over the output line. The observed pattern matched exactly that of the first 32 channels as presented in Fig. 4.6. This excludes the possibility that the higher noise in the channel numbers above 120 is the result of tail effects in the off-detector or DAQ electronics. Lab-measurements

[7]The performance of the Beetle was compared with that of another candidate for the vertex detector front-end chip, the SCTA chip. The same common mode strategy was applied for the analysis of the data from both chips.

Table 4.3: Results of the noise studies using 40k events. The standard deviation of the pedestal subtracted ADC-spectrum is given for the four different common-mode strategies. The column labeled "System noise" contains the average system noise as described in the text. The column "No correction" represents the results when no common-mode subtraction is applied. All values are in ADC counts.

Chip nr.	System noise	No correction	Corrected noise			
			1	2	3	4
0	5.1	19.3	16.6	16.4	16.5	16.3
1	6.1	20.7	17.7	17.5	17.6	17.3
2	5.5	19.0	16.3	16.2	16.2	16.0
3	6.1	22.0	19.5	19.3	19.2	19.1
4	6.9	21.3	18.7	18.6	18.6	18.3
5	6.0	20.1	16.9	16.7	16.8	16.5
6	6.8	21.3	17.0	16.8	16.9	16.6
7	8.1	19.3	16.7	16.5	16.6	16.4
8	5.5	21.3	16.8	16.6	16.7	16.5
9	7.2	19.4	16.8	16.6	16.7	16.5
10	4.7	21.3	16.0	15.9	15.9	15.8
11	5.8	22.7	20.0	19.8	19.8	19.5
12	5.9	20.8	18.7	18.6	18.6	18.4
13	5.5	18.4	16.1	15.9	16.0	15.8
14	5.3	17.9	15.8	15.7	15.7	15.5
15	10.6	17.7	15.9	15.8	15.8	15.7

will be done for future chip versions, to investigate whether this problem can be attributed to the inhomogeneity of the noise in the pipeline cells [60].

The data from the outliers mentioned in table 4.2 are rejected and for the remaining channels weight correlation factors can be calculated. These weight factors, which were also required to apply strategy 4, are shown in Fig. 4.7.

Here, the apparent structure shows up again: the first two channels in a group of 4 pick up more common-mode noise than the last two channels. The channels at the end of the chip pick up hardly any common-mode at all. When these corrections are applied in the common-mode subtraction, the noise of the channels with channel numbers above 120 is still higher than the noise of the rest of the channels. Although the weight factors typically deviate quite a lot from 1, this procedure yields on average only 1% extra reduction of the noise. This is due to the fact that the common-mode noise has an RMS of about 10 ADC counts; a 20% difference in the weight factor modifies the common-mode noise only by about 2 ADC counts. This correction is uncorrelated with the other noise sources, which contribute about 16 ADC counts.

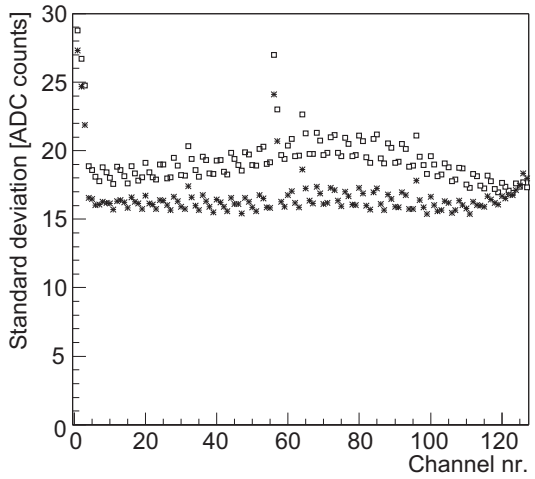


Figure 4.6: The total noise (open squares) and the common-mode subtracted noise (solid stars) in chip 0 as a function of channel number.

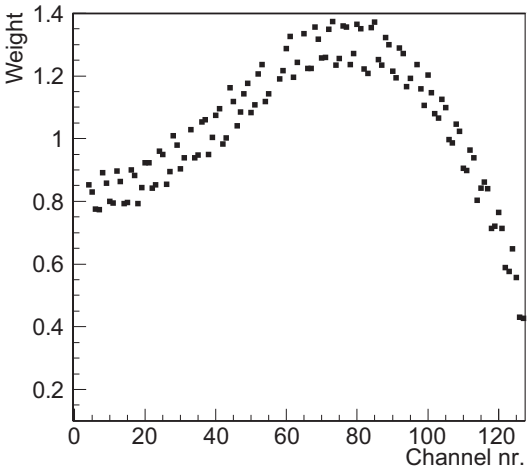


Figure 4.7: The weight factors as a function of channel number for the channel-dependent common-mode subtraction in strategy 4.

4.3 Pulse shape parametrization

The charge deposition of a particle that traversed the detector is analysed with three different methods which are described in section 4.4. The collected charge is expressed in units of signal/noise ratio to allow for easy comparison between the performance of various settings and chips. The signal/noise ratios are stored in a two dimensional histogram of amplitude versus time as shown in Fig. 4.8. A selected event gives 8 entries in this histogram each 25 ns apart. The data are binned in slices of 3 ns along the x-axis to reconstruct the transient response of the Beetle chip with sufficient time resolution.

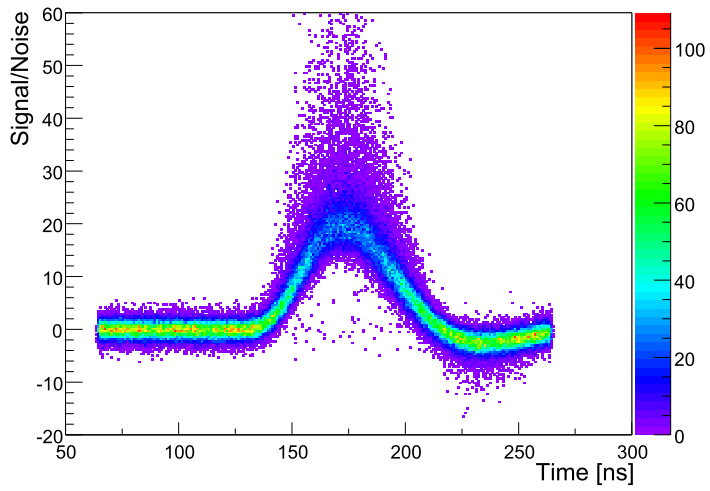


Figure 4.8: Two dimensional histogram of the signal amplitude (normalized to the width of the noise in the base line) versus time.

From each slice, a 1 dimensional histogram is constructed with the number of entries versus amplitude. These so-called energy-loss distributions are Gaussian in the baseline with a width of about 1. In the pulse they can be described by a Landau shape convoluted with a Gaussian. Details of this distribution are discussed in section 4.3.1. The characterization of the pulse shape is discussed in section 4.3.2.

4.3.1 Energy-loss distribution

The energy-loss distribution of a minimum ionizing particle passing through a thin slab of material is in first approximation described by the Landau probability density function [61]. Landau used a one-parameter model to describe the energy deposition of a traversing particle based on the properties and geometry of the stopping material. The binding energy of the electrons is neglected. As a consequence, the width of the distribution is underestimated. A practical solution is to broaden the Landau by means of a convolution with a Gauss [62]. In addition, the Gaussian convolution will account for noise contributions, which need to be disentangled when the results are compared to those of future measurements.

In total there are four free parameters: the most probable value (MPV) of the Landau, the width of the Landau, the width of the Gauss and a normalization parameter. The width and MPV of the Landau are decoupled for convenience[8]. The MPV of the convolution, which is quoted as signal/noise value, is a non-trivial combination of the fit parameters. This value is higher than the MPV of the Landau, its statistical error needs to be calculated by using the covariance matrix of the fit in combination with the sensitivity to the fit parameters [63]. In Fig. 4.9 a typical energy-loss distribution in the peak of the pulse is shown together with the fitted convolution of a Landau and a Gauss.

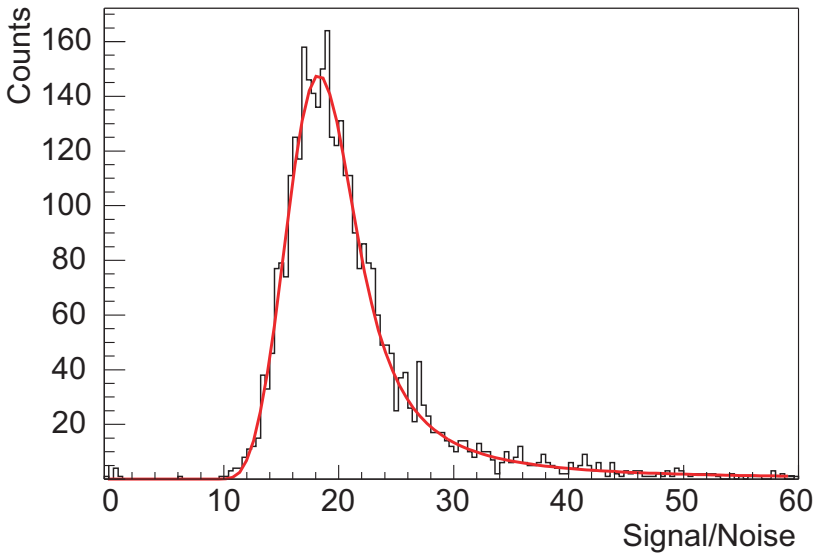


Figure 4.9: Typical energy-loss distribution in the peak of the pulse, fitted with the convolution of a Landau and Gauss.

[8] http://wwwasdoc.web.cern.ch/wwwasdoc/shortwrupsdir/g110/top.html

The convolution is performed with a numerical integration that uses Gaussian quadrature, based on an optimal distribution of the abscissa. This method is superior to trapezoid or Newton-Cotes integration methods which are based on equally spaced abscissa [64], especially in the rising edge and tail of the pulse where the width of the Landau becomes small.

4.3.2 Pulse shape characteristics

In Fig. 4.10 the MPV of each convolution is plotted versus the mean time of the slice it belongs to. From these kind of figures the pulse shape characteristics are extracted. The most interesting characteristics for *LHC*b are: rise time, signal/noise ratio and spill-over. Rise time is defined as the time difference between the moments where the pulse height amounts to 10% and 90% of the maximum pulse height. Signal/noise ratio is the maximum pulse height divided by the width of the noise in the baseline. Spill-over is the value of the pulse height 25 ns after the pulse reached its maximum divided by the maximum pulse height, *i.e.* in the *LHC*b experiment this fraction of the pulse will show up in the next bunch crossing. These parameters are extracted by using a 6^{th} order polynomial fit to the data in Fig. 4.10. This function gives a reasonable description of the data. The normalized χ^2 of this fit is 3.3 because of a slight mismatch in the rising edge and the peak of the pulse. The small over-estimate in the peak value is about 0.1 in signal/noise ratio, which is included in the systematic error. The pulse shape characteristics are extracted numerically. The statistical errors of these characteristics are determined in first order with the covariance matrix of the polynomial fit and the sensitivity to the fit parameters [63]. The pulse is divided in slices of 3 ns as mentioned in section 4.3. The effects of varying the width of the slices and the bin locations are found to be: $\pm$ 0.1 for the signal/noise ratio, $\pm$ 1% in absolute value for the spill-over, $\pm$ 0.5 ns for the rise time. These results are also included in the systematic error.

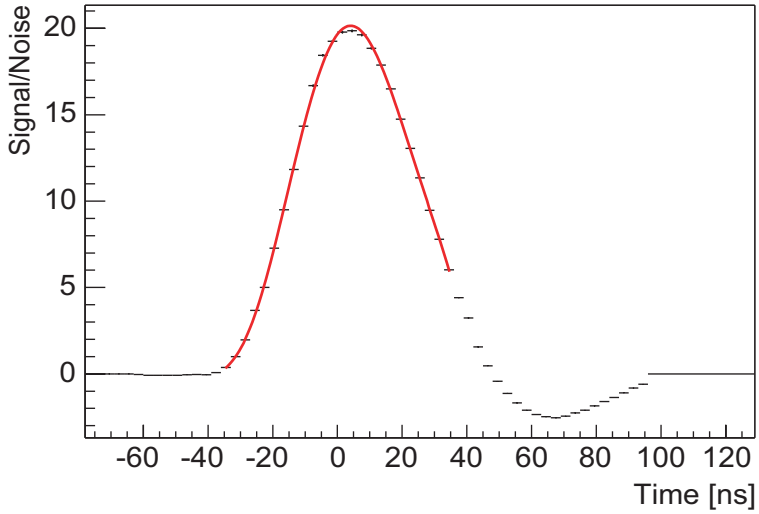


Figure 4.10: Most Probable Value (MPV) of the convolution plotted versus the mean time of the slice. The solid curve is a 6^{th} order polynomial fit through the MPVs. I_{sha} = 32 μA, I_{pre} = 350 μA and V_{fs} = 700 mV.

4.4 Analysis

The analysis tools make extensive use of the classes provided by the VeloROOT software[9] of the *LHC*b vertex detector group. The tools are customized for data format and detector geometry to be able to analyze the data produced by the Beetle chips and the XY station. The various histograms are created with a customized ROOT Tree[10] containing the events produced by the Beetles, extended with analysis methods to perform *e.g.* pedestal calculation, common-mode calculation and signal extraction. The preferred method to determine the Beetle pulse shape is via reconstruction of cluster charges as a function of time. This method, which is explained in section 4.4.1, is used for the vast majority of the data, because it generates the 'cleanest' distributions. The problem with this method is the risk of overestimating the performance. Adding the signal/noise ratio of a neighboring strip, only when its value is above a certain threshold can introduce a bias toward higher signal/noise ratios [65]. The results from our cluster charge analysis are compared to a strip charge analysis in section 4.4.2 and to a track selection analysis in section 4.4.3 to identify possible systematic effects. Furthermore, measurements were done to test the compliance with single time sample and with high trigger rate conditions, as required for *LHC*b operation (see section 4.4.4 and 4.4.5). In section 4.4.6 efficiency and spill-over as a function of threshold are discussed, allowing for a more direct estimate of the physics performance of the vertex detector if equipped with Beetle chips.

4.4.1 Cluster charge analysis

In the cluster charge analysis clusters are formed that represent the probability to find a hit at the reconstructed position. The signal/noise ratios of the selected strips are linearly added. This is mathematically equivalent to adding all the charge of a cluster and dividing it by the linear average of the noise in the strips that contribute to that cluster. Each event is treated in the following way:

1. Look for hits above 5σ, where σ is the noise of the strip after pedestal subtraction and common-mode noise correction.
2. Store the value of the highest hit divided by the noise of the corresponding strip for that time sample. This takes care of the normalization of the energy-loss distribution, which is why the signal amplitude in the histograms is expressed as signal/noise ratio.
3. If one or both of the adjacent strips have a value above 1σ, add the signal/noise ratio(s) to the cluster.
4. Histogram the signal/noise ratio of the cluster in the bin with the corresponding TDC value. Note that the content of the same channels is added for the other time samples, which gives seven additional entries each 25 ns apart in the 2 dimensional histogram.
5. Repeat steps 1-4 for the hits that have not yet been used in other clusters until all hits of this event are processed. If an event contains more than 4 clusters, it is considered noisy and discarded.

[9] http://lhcbproject.web.cern.ch/lhcbproject/velo/testbeam/software/doc/veloroot/html/index.html
[10] http://root.cern.ch

The result of this procedure is a plot like Fig. 4.8. Two slices of this plot, containing typical energy-loss distributions and fits, are shown in Fig. 4.11. Left the data from a 3 ns window around the peak and right from a 3 ns window 25 ns later (around the spill-over point). The distribution in the peak of the pulse resembles a Landau and 25 ns later it is mostly Gaussian. The normalized χ^2-values are 1.1 for both fits, indicating that the convolution gives a good description of the data.

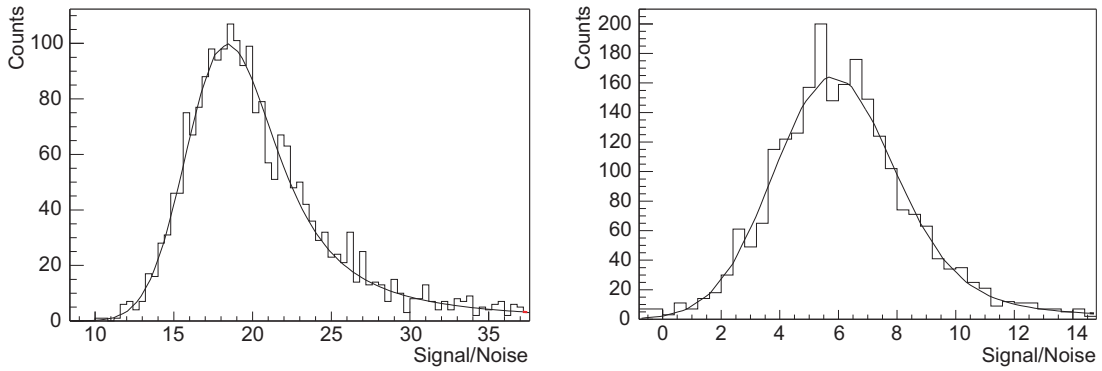


Figure 4.11: Energy-loss distributions for the cluster charge analysis of chip 8 with V_{fs} = 500 mV. On the left hand side the data from a 3 ns window around the peak are shown and on the right hand side from a 3 ns window 25 ns later (around the spill-over point).

The high statistics measurement for chips 8, 9, 10, 13, 14 and 15 (data set 2) enables the identification of trends within the chips. Sufficient data are available to determine the pulse shape characteristics for groups of twelve neighboring channels. The plots in Figs. 4.12 and 4.13 of peak amplitude, spill-over and peaking time differences[11] versus channel number, demonstrate clear trends as a function of channel number. Two groups of about 40 channels can be distinguished. The pulses in the higher-numbered channels are larger and slower than the ones in lower numbered-channels. In between these two groups there is a transition that spawns about 40 channels as well. The differences in pulse shape characteristics are visualized in Fig. 4.14 where the average pulse shape produced by channel numbers 5-45 is shown together with that produced by channel numbers 80-120.

The variations in the characteristics of the chip as function of channel number, as shown in Figure 4.12 and 4.13, hint to a problem in the chip. In fact, it relates to a problem that was first encountered in independent laboratory tests at both NIKHEF and Heidelberg: the first read-out after the one that contains a test pulse still contains a remainder of the test pulse, even when this read-out is delayed by 100 μs. This is explained by trapped charge in a capacitance of the read-out amplifier circuit of the Beetle chip (see Fig. 4.15) that transfers $\approx$-35% of the signal to the next read-out. This phenomenon has been baptized sticky charge [57]. The 8 consecutive samples used in the beam-test, together with the high statistics, make this problem visible in the data of the low-numbered chip channels. The resulting change of the pulse shape is illustrated schematically in Fig. 4.16. The analysis of the single time sample

[11] Peaking time is the time between the moments that the signal starts to rise and reaches its maximum. Only the difference in peaking time can be determined accurately because the start of the signal is difficult to detect.

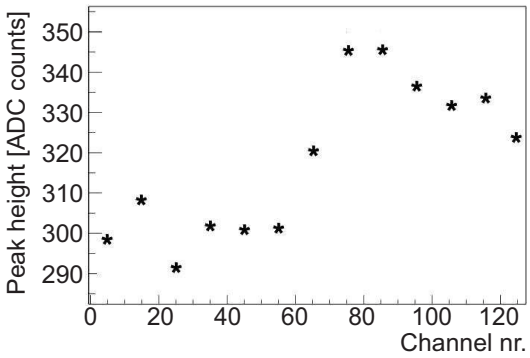


Figure 4.12: Peak amplitude versus channel number of chip 8.

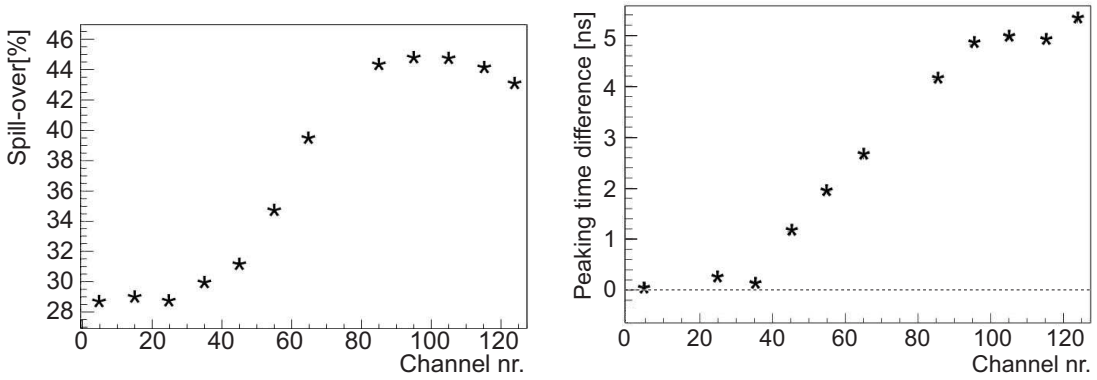


Figure 4.13: Spill-over (left) and peaking time differences (right) versus channel number of chip 8. The difference in peaking time is shown with respect to the first group of 12 channels.

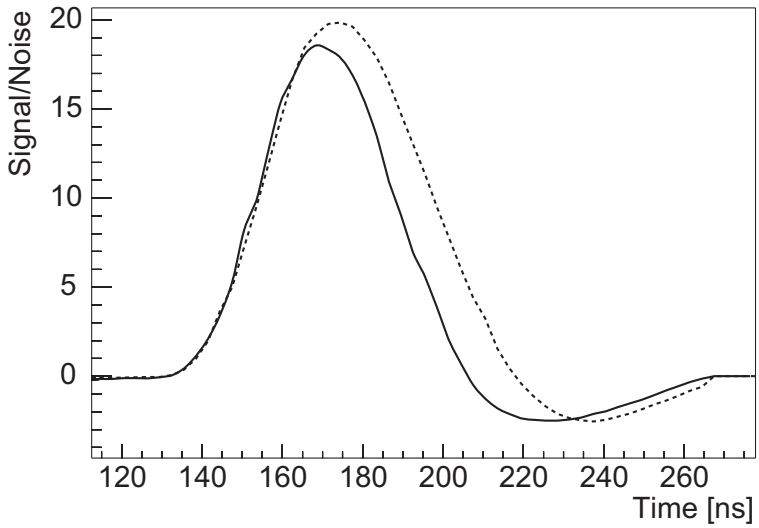


Figure 4.14: Pulse shapes for high (dashed curve) and low (solid curve) channel numbers of chip 8. The curves connect the datapoints.

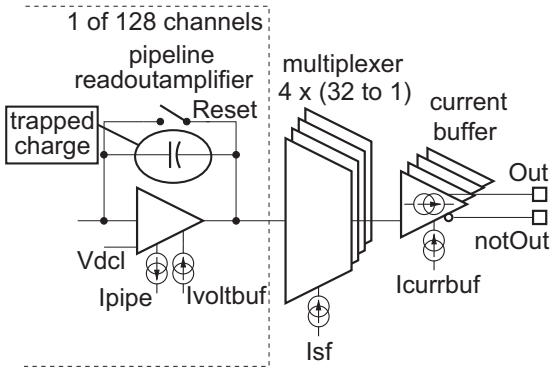


Figure 4.15: Read out circuit of the Beetle, for a functional diagram see Fig. 3.26. The capacitor with the trapped charge is indicated.

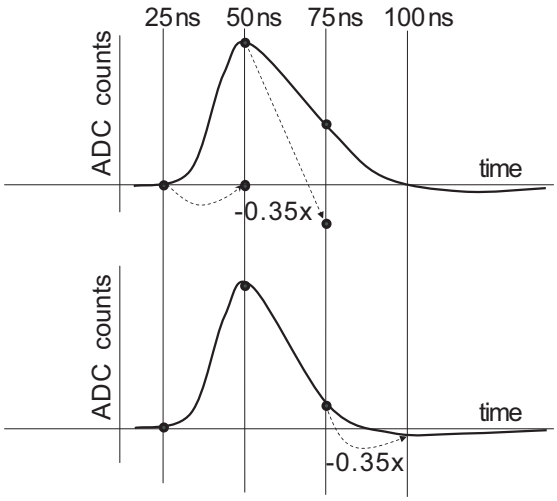


Figure 4.16: Sketch of the sticky charge effect: original pulse shape (top), affected pulse shape (bottom).

measurement in section 4.4.4 confirms a clear channel dependence and makes it possible to quantify the problem more accurately. Simulations indicate that this effect is time critical. It is caused by the simultaneous activation of the reset switch in the pipeline read out amplifier and a reset switch in the multiplexer. The problem disappears if the activation of one of them is delayed by about 1 ns[12]. Due to small delay differences within the chip only the channels with low channel numbers are affected. In case of consecutive triggers, the sticky charge problem is effectively a tail cancellation, because these samples are only 25 ns apart. The problem is solved in the Beetle1.2 by a forced discharge of the capacitance of the read-out amplifier in between read-outs[13]. This is a patch that only works for non-consecutive read-outs. In the Beetle1.3 the sticky charge problem is solved completely by changing the timing of the multiplexer circuit with respect to the read-out amplifier circuit by 5 ns, avoiding the simultaneous activation of the reset switches[14].

High-numbered channels do not suffer from this sticky charge effect, as will be shown in section 4.4.4 as well. Therefore, these channels can be used to make realistic predictions for the pulse shape of future chip versions. In the following analysis, all the numbers that are shown are based on the results of the channels with channel numbers 80-120. The channels 121-128 are excluded because the higher noise of these channels does not represent the average behavior of the chip (see Fig. 4.6).

The Beetles were programmed with various shaper settings to optimize the pulse shape for the vertex detector requirements. The dominant requirements are spill-over and signal/noise ratio. A setting was selected with 10-15% maximum undershoot. The shaper current was 32 μA and the pre-amplifier current was 350 μA. The feedback voltage of the shaper was varied to obtain 30% spill-over with the highest signal/noise ratio averaged over all channels. The results are presented in Table 4.4.

[12] http://wwwasic.kip.uni-heidelberg.de/lhcb/Publications/LHCb/LHCb_May2003_DB.pdf
[13] http://wwwasic.kip.uni-heidelberg.de/lhcb/Meetings/Meeting220103/Beetle_Lab_Measurements.pdf
[14] http://wwwasic.kip.uni-heidelberg.de/lhcb/Publications/LHCb/LHCb_Nov2003.pdf

Table 4.4: Overview of the average pulse shape characteristics for channel numbers 80-120. Signal/Noise is abbreviated with S/N, risetime with RT and spill over with SO (I_{sha} = 32 µA, I_{pre} = 350 µA).

	V_{fs} = 500 mV			V_{fs} = 700 mV			V_{fs} = 800 mV		
Chip nr.	S/N	RT [ns]	SO [%]	S/N	RT [ns]	SO [%]	S/N	RT [ns]	SO [%]
0	18.9±0.1	23.1±0.1	37.5±0.4	20.2±0.1	24.4±0.1	45.8±0.5	-	-	-
1	18.2±0.1	23.9±0.1	35.1±0.5	19.3±0.1	24.3±0.1	44.4±0.7	-	-	-
2	18.7±0.1	23.5±0.2	36.5±0.6	19.8±0.1	24.3±0.2	44.3±0.9	-	-	-
3	13.7±0.1	24.7±0.2	43.6±0.9	14.8±0.1	25.6±0.2	49.4±0.7	-	-	-
4	14.8±0.1	24.4±0.2	40.8±0.7	16.1±0.1	25.4±0.2	47.0±0.7	-	-	-
5	18.2±0.1	23.8±0.1	38.6±0.4	19.6±0.1	24.7±0.1	47.5±0.5	-	-	-
6*	16.2±0.1	24.4±0.2	35.9±0.6	16.7±0.1	24.3±0.2	43.5±0.7	-	-	-
7*	16.4±0.1	23.6±0.2	37.7±0.6	18.4±0.1	24.3±0.2	43.7±0.8	-	-	-
8	18.8±0.1	23.1±0.1	36.7±0.3	20.2±0.1	24.3±0.1	45.7±0.2	20.8±0.1	24.8±0.1	50.4±0.4
9	17.3±0.1	23.1±0.1	33.8±0.4	18.8±0.1	24.1±0.1	42.9±0.3	17.5±0.1	24.9±0.1	48.0±0.5
10	19.0±0.1	23.0±0.1	35.4±0.4	20.5±0.1	24.3±0.1	45.3±0.3	24.3±0.1	24.7±0.1	50.4±0.6
11	-	-	-	15.2±0.1	25.0±0.1	55.4±0.8	-	-	-
12	-	-	-	17.4±0.1	24.1±0.1	52.8±0.6	-	-	-
13	18.4±0.1	23.3±0.2	35.1±0.5	19.7±0.1	24.1±0.1	44.2±0.3	20.6±0.1	24.9±0.1	49.8±0.6
14	17.8±0.1	23.4±0.1	36.2±0.5	19.0±0.1	24.6±0.1	45.2±0.4	22.8±0.1	25.0±0.1	51.2±0.6
15*	17.2±0.1	23.4±0.2	34.6±0.6	18.6±0.1	23.9±0.1	44.0±0.4	22.7±0.1	24.3±0.1	48.0±0.6
*Higher system noise due to damaged DAQ channels									

Chips 3, 4, 11 and 12 are connected to the regions with the longer strips. The input capacitance of these chips is larger and their performance is inferior to the performance of the other chips [56]. The difference between the two kinds of tails of the hybrid might affect the performance as well. The tail constructed with the solid copper layer is used for the chips with numbers 0 - 7 and the tail with the grid for chips 8 - 15. On average the performance of the high numbered chips is slightly better than that of the low numbered chips, which could indicate that the average parasitic capacitance is reduced when a copper grid is used for the tail.

In some cases chips 6, 7 and 15 have a lower signal/noise ratio because of increased system noise in the ADC. This is a problem related to ground loop currents and spontaneously appeared and disappeared during the measurements. It was known that data of chips 7 and 15, which were connected to ADC input 7, suffer from this. But it became apparent that chip 6, which was connected to ADC input 5, had similar problems. Note the anomalous signal/noise ratios for chips 9 (low) and 10 (high) for the V_{fs} = 800 mV setting. This is not understood and should be verified in future beam-tests.

The measured pulse shape for chip 8, one of the chips connected to a region with shorter strips, is visualized in Fig. 4.17 and the corresponding simulated pulse shapes are shown in Fig. 4.18. These simulations are based on realistic models of the chip internals. First, the electrical circuits of the Beetle1.1 were designed with Cadence [66]. Then a SPICE netlist was extracted from the layout that included the parasitic capacitances in the front-end circuit. This information was used in HSpice, a numerical analysis package from METASOFT, to determine the expected time response. The accuracy of the time response of this kind of simulations is expected to be about 10%.

The average results for the chips connected to the regions with the short strips are summarized and compared to simulations in Table 4.5. The statistical error is the RMS value of the variations from chip to chip. The values for the signal/noise ratio found in the simulation are normalized to the measurements with V_{fs} = 800 mV. The signal/noise ratios for the other settings of the simulation are extrapolated based on the assumption that the noise does not change[15]. The simulations and measurements are in reasonable agreement. According to the simulations, faster pulses can be created by a further reduction of V_{fs} at the cost of a lower signal/noise ratio.

This section is concluded with a summary of the systematic errors of the cluster charge analysis. The contributions are added linearly. The effects have all been discussed in section 4.3.2 except for a correction of the signal/noise ratio based on the results of the track selection analysis, which is discussed in section 4.4.3. The systematic error of the signal/noise ratio includes - 0.1 from the fit of the pulse shape, $\pm$ 0.1 from the bin size variations and - 0.7 from the track selection analysis. In total this adds up to an asymmetric error of (-0.9, +0.1). The systematic errors of the rise time and the spill-over only consist of contributions from bin size variations and are $\pm$ 0.5 ns and $\pm$ 1% (absolute value), respectively.

[15]This assumption is not completely correct: an increase of V_{fs} reduces the bandwidth of the shaper, which results in a reduction of the noise.

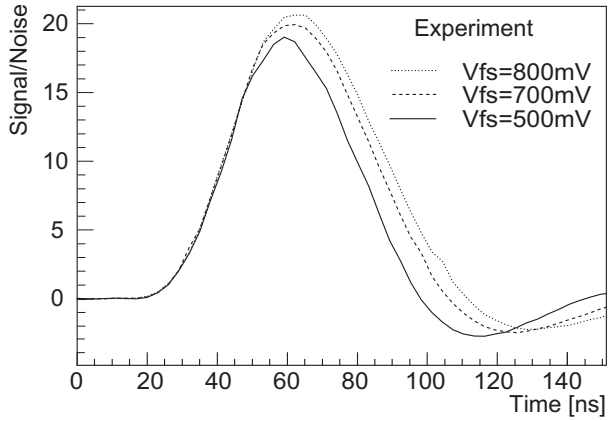


Figure 4.17: Overview of the MPVs of chip 8 for various bias settings. I_{sha} = 32 µA, I_{pre} = 350 µA, V_{fs} = 500 mV (solid curve), 700 mV (dashed curve) and 800 mV (dotted curve). The curves connect the data points.

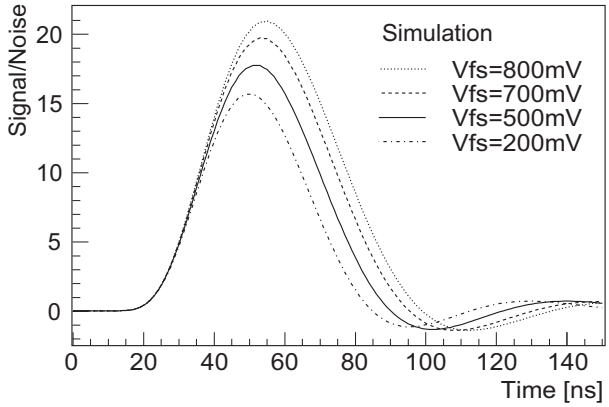


Figure 4.18: Overview of the pulse shape simulations for various bias settings. I_{sha} = 32 µA, I_{pre} = 350 µA, V_{fs} = 200 mV (dash-dotted curve), 500 mV (solid curve), 700 mV (dashed curve) and 800 mV (dotted curve).

Table 4.5: Comparison of simulations with averages of the cluster charge analysis for various V_{fs} settings (I_{sha} = 32 µA, I_{pre} = 350 µA). The signal/noise ratio of the simulation is normalized at V_{fs} = 800 mV and extrapolated to the other settings based on the assumption that the noise is constant. The statistical error of the analysis is the RMS value of the variations from chip to chip.

	Signal/Noise		Rise time [ns]		Spill-over [%]	
	Simu.	Experiment	Simu.	Experiment	Simu.	Experiment
V_{fs} = 500 mV	17.7	17.9±0.9	20.3	23.5±0.7	30	36.1±1.3
V_{fs} = 700 mV	19.7	19.2±1.0	20.7	24.3±0.2	35	44.7±1.3
V_{fs} = 800 mV	21.4	21.4±2.1	21.8	24.7±0.2	37	49.6±1.1

4.4.2 Strip charge analysis

The pulse shapes of signals from single strips are studied in what we will call the 'strip charge analysis'. For each event the strip with the highest signal/noise ratio is selected. For this strip the common-mode corrected signal/noise ratios of the 8 time samples are stored in a two dimensional histogram. The strip charge method is the simplest way to analyze the data. No knowledge is required about detector geometry, alignment, pitch adapter, *etc*. A disadvantage of this method is that only a fraction of the charge is found in case the charge is distributed over more than one strip. This results in broadening of the energy-loss distribution and a shift to a lower value.

In Fig. 4.19 the signal distributions in a 3 ns window around the peak and 25 ns later (around the spill-over point) are presented for chip 8 with V_{fs} = 700 mV. The fits have χ^2-values of 1.1, which indicates that they give a good description of the data.

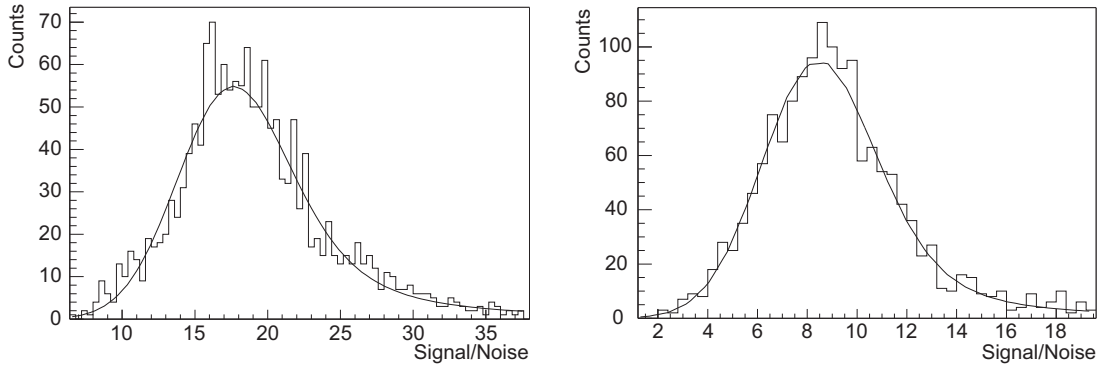


Figure 4.19: Energy-loss distributions and fits for the strip charge analysis of chip 8 with V_{fs} = 700 mV. On the left hand side the data from a 3 ns window around the peak are shown and on the right hand side from a 3 ns window 25 ns later (around the spill-over point).

In Table 4.6 the results of the strip charge analysis are compared to the cluster charge analysis for V_{fs} = 500 mV. The rise time and spill-over are almost identical, but the signal/noise ratio is 5 - 10% lower for the strip charge analysis. This is in accordance with the findings of Ref. [65] for a setup with signal/noise ratios comparable to ours. In the rest of this section we will investigate whether the lower signal/noise ratios found with this method are due to the charge sharing between neighboring strips. This effect is described by the parameter η, which represents the fraction of the cluster charge on the strip with the highest signal/noise ratio.

In Fig. 4.20 the measured distribution of η is shown, in which the data with $\eta > 0.9$ are collected in the bin at $\eta = 1$ because in this region the charge on the second strip can not always be distinguished from the noise. In the same figure a parametrization of the η-distribution is shown normalized to the number of entries with $\eta > 0.9$, which matches the data reasonably well.

The parametrization was calculated from a relation between η and the distance of the hit to the nearest strip, assuming that the incident particles are homogeneously distributed over the detector surface and that the charge is shared between at most 2 strips. This more intuitive relation is shown in Fig. 4.21. Due to geometrical arguments it is symmetric around

the point where the distance is half the strip pitch, which lies in the middle between the two strips.

Table 4.6: Comparison of the strip charge and cluster charge analysis (I_{sha} = 32 µA, I_{pre} = 350 µA, V_{fs} = 500 mV).

	Strip charge analysis					
Chip	Signal/Noise		Rise time [ns]		Spill-over [%]	
0	17.9±0.1		24.6±0.1		35.6±0.5	
1	16.9±0.1		23.5±0.1		37.2±0.5	
2	17.3±0.1		23.8±0.2		35.6±0.7	
3	11.8±0.1		24.6±0.2		44.2±0.8	
4	13.3±0.1		24.5±0.2		40.6±0.7	
5	17.4±0.1		24.1±0.1		38.2±0.4	
6	15.3±0.1		23.9±0.1		37.5±0.5	
7*	15.1±0.1		24.0±0.2		38.2±0.7	
8	17.7±0.1		23.4±0.1		36.9±0.4	
9	16.4±0.1		23.2±0.1		35.2±0.5	
10	17.1±0.1		23.3±0.1		36.6±0.5	
13	17.2±0.1		23.5±0.1		35.4±0.5	
14	16.9±0.1		23.6±0.1		37.4±0.6	
15*	15.5±0.1		23.5±0.2		35.8±0.7	
	Cluster charge analysis					
Chip	S/N	$\frac{(S/N)_{Cluster}}{(S/N)_{Strip}}$	RT [ns]	$\frac{RT_{Cluster}}{RT_{Strip}}$	SO [%]	$\frac{SO_{Cluster}}{SO_{Strip}}$
0	18.9±0.1	1.06	23.1±0.1	1.06	37.5±0.4	0.95
1	18.2±0.1	1.08	23.9±0.1	0.98	35.1±0.5	1.06
2	18.7±0.1	1.08	23.5±0.2	1.01	36.5±0.6	0.98
3	13.7±0.1	1.16	24.7±0.2	1.00	43.6±0.9	1.01
4	14.8±0.1	1.11	24.4±0.2	1.00	40.8±0.7	1.00
5	18.2±0.1	1.05	23.8±0.1	1.01	38.6±0.4	0.99
6	16.2±0.1	1.06	24.4±0.2	0.98	35.9±0.6	1.04
7*	16.4±0.1	1.09	23.6±0.2	1.02	37.7±0.6	1.02
8	18.8±0.1	1.06	23.1±0.1	1.01	36.7±0.3	1.01
9	17.3±0.1	1.05	23.1±0.1	1.00	33.8±0.4	1.04
10	19.0±0.1	1.11	23.0±0.1	1.01	35.4±0.4	1.03
13	18.4±0.1	1.07	23.3±0.2	1.01	35.1±0.5	1.01
14	17.8±0.1	1.05	23.4±0.1	1.01	36.2±0.5	1.03
15*	17.2±0.1	1.11	23.4±0.2	1.00	34.6±0.6	1.03
		1.08±0.03		1.01±0.02		1.01±0.03
*Higher system noise due to damaged DAQ channels. No data available for chips 11 and 12.						

Note that the width of the region with η = 1 in Fig. 4.21 matches half the strip width, which means that a hit directly on a strip is only supposed to deposit charge on that strip. A

hit in the middle of two strips gives an η of 0.5. The distribution of the number of strips that contribute to a cluster is extracted from the data and shown in Fig. 4.22. In 92% of the cases the cluster size is one or two. For the other 8% a (most of the times small) error is introduced corresponding to the charge that is not deposited on the two strips nearest to the hit.

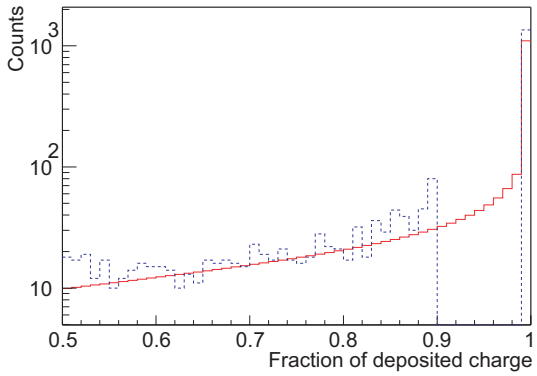


Figure 4.20: The η-distribution. The dashed histogram is calculated from the data and the solid one is a parametrization.

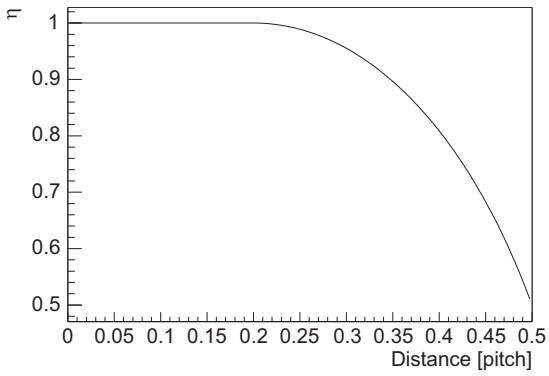


Figure 4.21: Relation between η and the distance of the hit to the nearest strip. The distance is expressed in units of pitch.

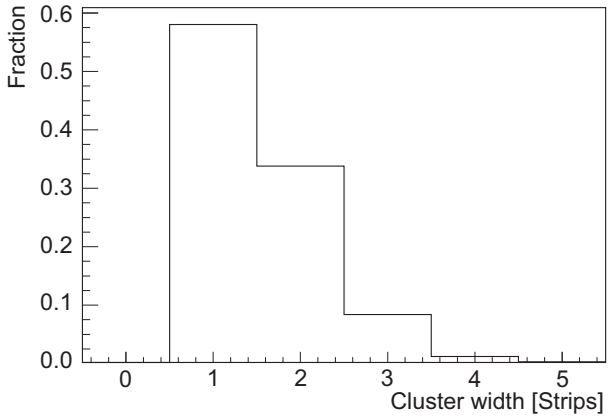


Figure 4.22: Number of strips that contribute to a cluster.

The parametrized η-distribution is used to calculate an expected result for the strip charge analysis from the cluster charge analysis. The entries in the histogram of the energy-loss distribution of the cluster charge analysis are redistributed with a folding procedure. A large fraction of the entries are folded with $\eta = 1$ and thus remain in the same bin. The remaining events move to lower signal/noise ratios according to the η-distribution. The result of this redistribution is compared to the histogram of the strip charge analysis in Fig. 4.23.

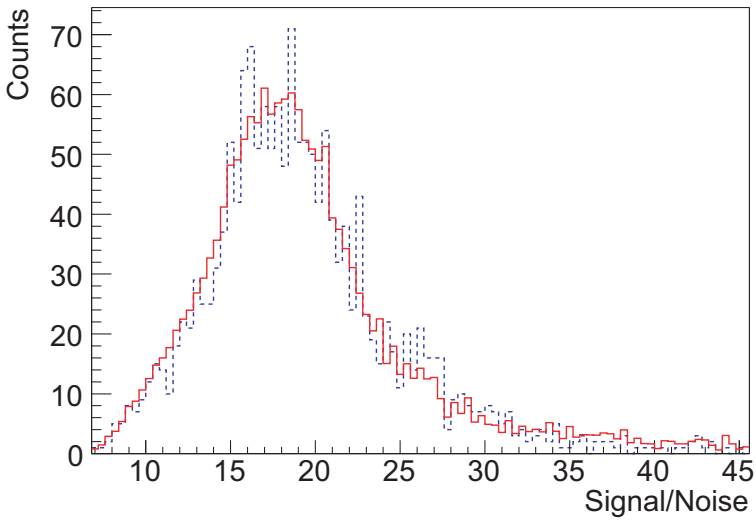


Figure 4.23: Comparison of the energy-loss distribution from the strip charge (dashed) analysis and the redistributed cluster charge (solid) using the η*-distribution.*

The two distributions agree very well. The distributions are fitted with a convolution of a Gauss and a Landau. All fit parameters match within 1.5σ, where σ is the error of the parameter. The MPVs, which are the key parameters for the signal/noise ratios mentioned in Table 4.6, are equal within 0.5σ. We conclude that the differences found between the results from the strip charge analysis and the cluster charge analysis can solely be attributed to charge sharing between strips.

4.4.3 Track selection analysis

In the 'track selection analysis' three strips nearest to the x and y coordinates found in an independent XY tracking station are selected[16]. For each of the eight time samples the signal/noise ratios of these strips are added and stored in a two dimensional histogram of amplitude versus time.

An advantage of the track selection method is that there is no a priori selection of strips based on signal/noise ratio, while the strip selection in the cluster charge method is entirely based on signal/noise ratio. A disadvantage of the 'track selection analysis' is that detailed knowledge is required of the detector bonding scheme, the detector strip geometry and the detector alignment with respect to the tracking station. A second disadvantage is that the amplitude is always the sum of the signal/noise ratios of three strips even if some of these strips carry no charge; these strips will only contribute to the width of the distribution. A third disadvantage is the dependence of the results on the quality of the track information. The strips that are selected may contain only a fraction of the charge or just noise because the track information is not sufficiently accurate. In Fig. 4.24 the distribution of the residuals between the R coordinate predicted by the tracking station and the cluster center in the Beetle

[16]The selection is based on the assumption that the tracks are perpendicular to the sensor plane. An independent measurement showed that the dispersion of the beam introduces an uncorrelated contribution to the resolution with a standard deviation of about 5 μm.

detector is shown. The fitted Gaussian has a standard deviation of about 30 μm (varying slightly with run conditions). The data for the track selection analysis have been collected on the inner region of the Beetle detector, where the strip pitch is 32.5 μm.

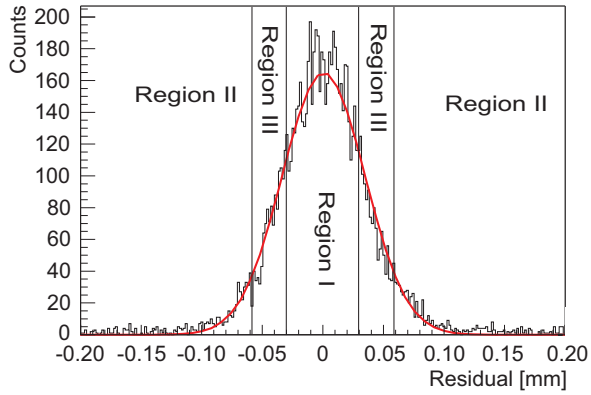


Figure 4.24: Residuals of the reconstructed tracks. Three regions can be distinguished. The track selection results in the reconstruction of all signal charge when the residual resides in region I. When the residual lies in region II, no signal charge is collected (just noise). In region III only part of the signal charge is reconstructed (the reconstructed fraction can be determined with the η*- function). These regions are discussed later in this section.*

In principle, the resolution of the tracking station should be much smaller than the strip pitch to allow for accurate reconstruction of a cluster with the track selection method[17]. However, the data can still be used to determine the pulse shape characteristics and to compare to the results of the cluster charge analysis. In Fig. 4.25 the energy-loss distributions in a 3 ns window around the peak and 25 ns later (around the spill-over point) are presented for chip 0 with V_{fs} = 500 mV. In addition to the expected Landau distribution, a clearly separated peak around zero is apparent in the first histogram, which is the result of insufficient track resolution. This pollution is approximated with a Gaussian and subtracted from all slices[18]. Note that in the second histogram of Fig. 4.25 the noise contribution can no longer be separated. The corrected distributions are shown in Fig. 4.26.

In spite of the correction, the histogram of the 3 ns window around the spill-over point still has a much larger width than the histogram of Fig. 4.11, created with the cluster charge analysis. This may be explained by the selection of 3 strips, which should broaden the distribution[19].

The results of the track selection analysis are compared to the cluster charge analysis in Table 4.7 for V_{fs} = 500 mV. Only those results are given where sufficiently accurate track

[17]The best resolution obtained with this XY detector is 8 μm [67]. This resolution was only obtained under conditions with low noise levels. In our setup additional noise was introduced by the electronics that was used to match the DC-levels of the signals with that of the ADC.

[18]The Gaussian around zero has a slightly positive mean value and is broader than expected from the noise. This indicates that strips further away than two times the pitch can still receive some charge. This is also observed in Fig. 4.22. The width and the position of the Gaussian are scaled linearly with an estimate of the MPV to compensate for this effect.

[19]Hits that are almost 2 strips away give a small signal. In the tail of the pulse these hits can not be separated from the main Landau and broaden the shape as well.

information and a sufficiently low noise background resulted in an acceptable normalized χ^2 for the fits to the energy-loss distributions. The fits for chips 0, 5 and 13 give the best normalized χ^2, thus these correspond to the most reliable results.

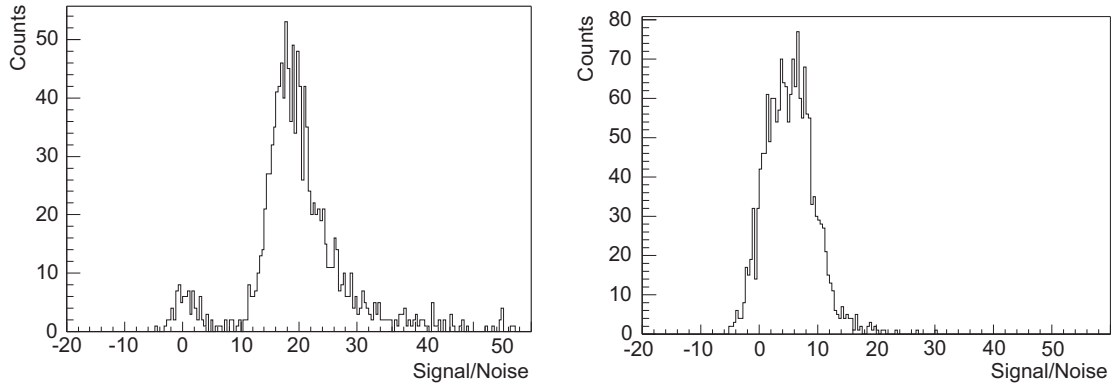


Figure 4.25: Energy-loss distributions for the track selection analysis of chip 0 with V_{fs} = 500 mV. On the left hand side the data from a 3 ns window around the peak are shown and on the right hand side from a 3 ns window 25 ns later.

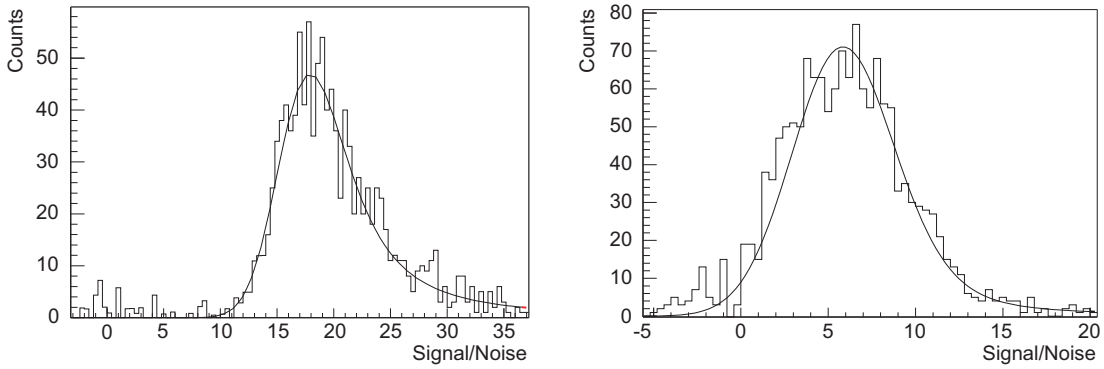


Figure 4.26: Energy-loss distributions for the track selection analysis of chip 0, corrected for pollution with noise, V_{fs} = 500 mV. On the left hand side the data from a 3 ns window around the peak are shown and on the right hand side from a 3 ns window 25 ns later.

The track selection analysis results in almost identical values for the rise time and in significantly lower values for the signal/noise ratio and the spill-over. The systematic uncertainty in the spill-over is difficult to estimate, but a conservative approach is to assume that the lower spill-over is completely due to systematic effects caused by insufficient resolution in the track selection analysis.

In the rest of this section the observed differences in signal/noise ratio are discussed. An expected result for the track selection method is determined from the energy-loss distribution obtained with the cluster charge analysis. Entries in the energy-loss histogram are redistributed based on the measured residual distribution shown in Fig. 4.24. When the distance between the *R* coordinate predicted by the tracking station and the cluster center is less than one unit of strip pitch, the signal/noise ratio is not changed, because the signal/noise ratios of the 3 strips nearest to the predicted location are added together and will on average give

the same signal/noise ratio as the cluster. Most of the hits will be in this range (region I in Fig. 4.24). When the residuals are larger than 2 units of strip pitch (regions II), the signal/noise ratios of the three strips will on average add up to zero. Thus a small fraction of the entries will be redistributed to a Gaussian peak around zero with a standard deviation of $\sqrt{3}$. When the residuals are between 1 and 2 units of strip pitch (regions III in Fig. 4.24) only a fraction of the charge will be found on the three strips around the predicted position. The η-distribution is calculated by combining the relation between η and the distance to the nearest strip as shown in Fig. 4.21 with distribution of the residuals (Fig. 4.24)[20]. The entries in the energy-loss distribution are transfered to lower signal/noise ratios according to this distribution.

Table 4.7: Comparison of the track selection and the cluster charge analysis (I_{sha} = 32 μA, I_{pre} = 350 μA, V_{fs} = 500 mV).

	Track selection analysis					
Chip	Signal/Noise		Rise time [ns]		Spill-over [%]	
0	18.3±0.1		23.1±0.1		32.5±0.6	
1	18.5±0.1		23.7±0.3		Not available	
2	18.0±0.1		23.7±0.3		26.6±0.9	
5	17.5±0.1		23.8±0.2		32.0±0.7	
6*	16.6±0.1		20.4±0.3		Not available	
7*	16.9±0.1		22.7±0.2		26.3±0.9	
13	17.6±0.1		22.8±0.2		29.9±0.8	
	Cluster charge analysis					
Chip	S/N	$\frac{(S/N)_{Cluster}}{(S/N)_{Track}}$	RT [ns]	$\frac{RT_{Cluster}}{RT_{Track}}$	SO [%]	$\frac{SO_{Cluster}}{SO_{Track}}$
0	18.9±0.1	1.03	23.1±0.1	1.00	37.5±0.4	1.15
1	18.2±0.1	0.98	23.9±0.1	1.01	35.1±0.5	-
2	18.7±0.1	1.04	23.5±0.2	0.99	36.5±0.6	1.37
5	18.2±0.1	1.04	23.8±0.1	1.00	38.6±0.4	1.21
6*	16.2±0.1	0.98	24.4±0.2	1.20	35.9±0.6	-
7*	16.4±0.1	0.97	23.6±0.2	1.04	37.7±0.6	1.43
13	18.4±0.1	1.05	23.3±0.2	1.02	35.1±0.5	1.17
		1.01±0.01		1.04±0.07		1.27±0.13
Chip 0, 5 and 13 only		1.04±0.01		1.01±0.01		1.18±0.03
*Higher system noise due to damaged DAQ channels.						
The statistics or data quality of the chips that are not shown						
was insufficient for analysis.						

In Fig. 4.27 the result of this redistribution of the energy-loss distribution obtained with the cluster charge analysis is compared to the result of the actual track selection analysis.

[20]The pitch on the x-axis of the η-distribution now ranges from 0 to 1 (instead of 0 to 0.5). Due to the convolution with this distribution a continuous background appears in the histogram between the noise around zero and the peak in the energy-loss spectrum around 20.

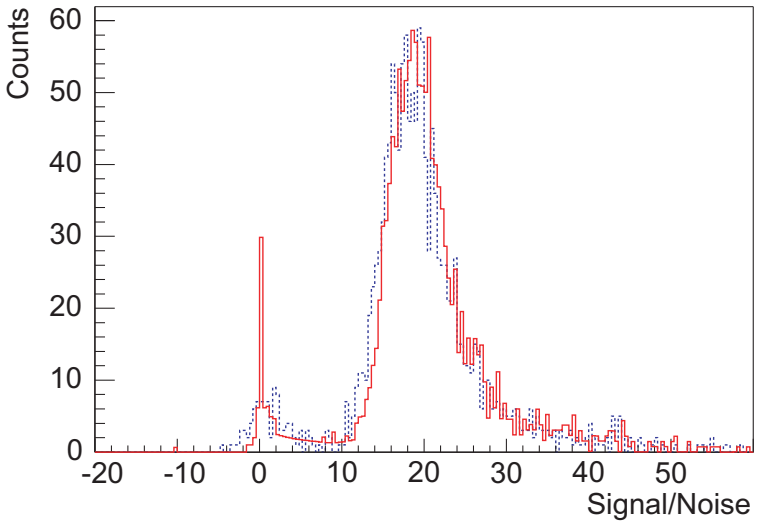


Figure 4.27: Comparison of the energy-loss distribution for the track selection (dashed) and the cluster charge analysis, using the η-distribution and the resolution (solid).

A small shift can be observed between the distribution obtained in the track selection method and the redistribution of the cluster analysis. The histograms are fitted with a convolution of a Landau and a Gauss. All the fit parameters match within 1.5σ (where σ is the error of the parameter) except for the MPVs, which are the key parameters for the signal/noise ratios. For the track selection analysis the MPV is about 0.7 lower than for the cluster analysis, which matches the differences found for chip 0, 5 and 13 in Table 4.7. Although the average signal/noise ratio is reduced by the redistribution, it does not significantly change the MPV. Hence, the small shift can not be explained by the finite resolution of the tracking station in combination with the η-distribution[21]. A possible bias in the cluster charge method is the choice of the cluster size, which is based on thresholds and selects only positive contributions to the signal/noise ratio. As the discrepancy can be an indication of a systematic effect of the cluster charge method, the observed difference is included in the systematic error.

4.4.4 Single time sample measurements

In the previous sections all data were acquired for 8 consecutive time samples. This allows the study of the complete pulse shape without changing the latency. However, in the actual *LHC*b running conditions, data will only be collected in a single time sample. Therefore, we also collected so-called single time sample data to investigate whether or not the performance of the Beetle depends on the number of consecutive samples. Independent runs were taken for 25 ns regions enclosing the peak, the rising edge and the tail of the pulse shape.

Single time sample data were taken for chips 0, 1, 2 , 5, 6 and 7 with V_{fs} = 700 mV. Track information is needed for the strip selection, in particular for an accurate reconstruction of the rising and falling edge of the pulse. Chip 0 is selected for this analysis because

[21] The sensitivity of this result to the resolution and to the exact shape of the η-distribution is small. The MPV is dominated by the events with fully reconstructed charge, both in the tracking and in the cluster charge analysis.

it has the highest statistics and the most accurate track reconstruction. The results are compared to those of the cluster charge analysis in Table 4.8, where the systematic errors of the cluster charge analysis are also listed. The results are in good agreement; no degradation in performance is observed if only one time sample of the Beetle is read-out.

Table 4.8: Comparison of the single time sample and the cluster charge analysis for chip 0 (I_{sha} = 32 μA, I_{pre} = 350 μA, V_{fs} = 700 mV). The systematic errors of the cluster charge analysis are listed.

	Signal/Noise	Rise time [ns]	Spill-over [%]
Cluster charge analysis	$20.2\pm0.1^{+0.1}_{-0.9}$	24.4±0.1±0.5	45.8±0.5±1
Single time sample analysis	19.6±0.1	25.0±0.3	45.6±0.5

Single time sample data are also used to get more accurate information about the sticky charge effect. The time difference between two successive triggers of the Beetle was about 10 ms . Therefore, the sticky charge effect cannot be caused by the pulse shaping circuitry of the front-end. This possibility could not be excluded from the data with multiple time samples. Events are selected with a signal above 5σ. For these events, the magnitude of the sticky charge effect is the ratio of the signals in the next and current event. The average results are shown in Fig. 4.28, the vertical bars indicate the standard deviations of the distributions. The figure shows that the sticky charge effect in the Beetle1.1 is large and negative for channel numbers below 40 and absent for channel numbers above 80. In the Beetle1.3 this problem is solved.

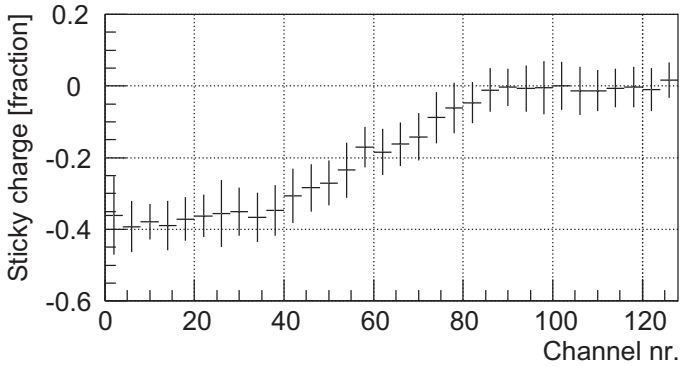


Figure 4.28: Sticky charge ratio as function of channel number.

4.4.5 High trigger rate analysis

The high trigger rate test was introduced to operate the chip under conditions where the read-out is almost continuously active. In that case, an incoming signal from the detector is processed during an ongoing read-out. Additional noise may then be generated due to interference of the sampling and the read-out process. Test pulse triggers, each issuing the read-out of 4 consecutive time samples, were generated at a rate of 40 kHz. This occupied the Beetle read-out circuit for 64% of the time. These test pulse triggers were sent only to the Beetle, so the corresponding data were not acquired. In conjunction with the test pulse

trigger, the physics trigger which was generated by the scintillators was passed both to the Beetle and the ADC, so these physics data were read-out and stored. The measurements were taken with V_{fs} = 700 mV.

Unfortunately, a large fraction of the data is polluted with test pulses due to malfunctioning electronics. The data are filtered and collected in one histogram containing signal/noise ratios from the channels 80-120 of all available chips (0, 1, 2, 5, 6 and 7). The results are compared in Table 4.9 to the results from a combined data set of the cluster charge analysis with the same chips. Except for the signal/noise ratio, the numbers are in good agreement with each other. Although the average noise is slightly higher in case of the high trigger rate data, this does not account for the large difference between the two signal/noise ratios. The decrease is caused by a lower signal in the data from the high trigger rate measurements compared to that of the other data. This may be attributed to interference of the test pulse circuit with the read-out, but the mechanism is yet unclear. More measurements are required to draw definite conclusions on this issue.

Table 4.9: Comparison of the high trigger rate and the cluster charge analysis (I_{sha} = 32 µA, I_{pre} = 350 µA, V_{fs} = 700 mV). Combined result of chips 0, 1, 2, 5, 6 and 7.

	Signal/Noise	Rise time [ns]	Spill-over [%]
Cluster charge analysis	19.2±0.1	24.3±0.1	44.7±0.1
High trigger rate	14.6±0.1	24.7±0.1	43.5±0.5

4.4.6 Efficiency, noise and spill-over

Although the pulse shape characteristics are well suited to compare the performance of various settings and chips, they are not intuitively connected to the physics performance. A better approach is to show the efficiency of the Beetle detector versus the threshold in units of signal/noise ratio. This is done in Fig. 4.29 for region 0 of the sensor. The single time sample data with V_{fs} = 700 mV are used. The corresponding pulse shape characteristics are: signal/noise ratio = 19.6, rise time = 25.0 ns and spill-over = 45.6 %. The curve is generated by scanning the threshold from a signal/noise ratio of 5 to 45. The following steps are taken to create this figure.

1. Select events with a hit in the selected region of the tracking station.
2. Try to match each hit found in the tracking station with a cluster above threshold in the Beetle detector within a window of 5 ns around the peak.
3. When all events are processed, the efficiency is equal to the number of clusters found in the Beetle detector divided by the number of selected events in the tracking station.

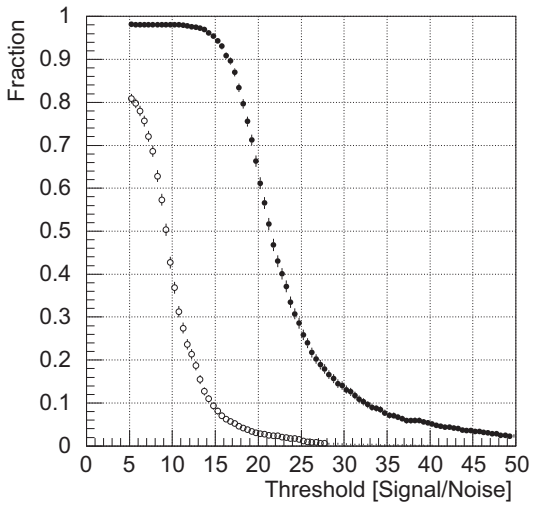


Figure 4.29: Efficiency (filled circles) and fraction of spill-over hits (open circles) as a function of threshold for chip 0. The time window for accepted hits in the Beetle detector is 5 ns.

The curve starts at a threshold of 5σ to avoid pollution of the plot with noise. For low thresholds almost all hits in the tracking station can be matched with a corresponding cluster in the Beetle detector. The few percent inefficiency can be due to either a noise hit in the tracking station, a broken strip or channel in the Beetle detector, or a non-perpendicular track.

In the *LHC*b experiment the fraction of hits above threshold after 25 ns will pollute the physics data, since remnants of the hits will still be visible in the next bunch crossing. Therefore, the fraction of hits caused by spill-over is calculated and shown as well. This curve is generated in the same way as the efficiency curve, however the sampling window for the Beetle detector is delayed by 25 ns.

The efficiency will almost be 100% when the threshold is set at a signal/noise ratio of 12, but about 20 to 25% of these hits will also be seen in the next bunch crossing. Raising the threshold to 15 reduces this contribution to 10% at the cost of about 3% efficiency-loss[22].

4.5 Performance summary

A prototype vertex detector hybrid was successfully designed, constructed and tested in a beam environment. The 16 Beetle1.1 chips were bonded without prior testing. They are all fully operational, which indicates a high yield of the associated production run in deep submicron technology. The beam-test experiment showed no abnormal behavior of the chips due to their mounting closely together on a hybrid.

The majority of the data is analyzed with a method that uses cluster charges. This method is cross-checked both with a method that uses the signal/noise ratio of a single-strip and with a method based on track selection. The comparison with the track selection analysis indicates a possible overestimate of about 0.7 in signal/noise ratio for the cluster charge analysis. This is included in the systematic error. Other small systematic uncertainties are due to the finite

[22]Note that the optimal threshold setting for *LHC*b also depends on the tuning of the pulse shape.

bin size and a slight tendency of the fit to overestimate the maximum value of the pulse. The pulse shape characteristics of the setting that complies best with LHCb specifications (I_{sha} = 32 μA, I_{pre} = 350 μA, V_{fs} = 500 mV) for the regions with the short strips are shown in Table 4.10 together with the results from simulation. If a slightly faster pulse is desired, the parameters can be re-tuned at the cost of a somewhat lower signal/noise ratio.

Table 4.10: Comparison of simulation with the average of the cluster charge analysis (I_{sha} = 32 μA, I_{pre} = 350 μA, V_{fs} = 500 mV).

	Signal/Noise	Rise time [ns]	Spill-over [%]
Analyzed	$17.9\pm0.9\,^{+0.1}_{-0.9}$	$23.5\pm0.7\pm0.5$	$36.1\pm1.3\pm1$
Simulated	n.a.	20.3	30

Single time sample data and high trigger rate data are analyzed and found to be in good agreement with the rest of the results, except for the low signal/noise ratio of 14.6 of the high trigger rate analysis compared to 19.2 for the cluster charge analysis. This could be caused by interference from the test pulse circuit, but more measurements are required to draw definite conclusions on this issue. No other deteriorating effects are found due to these two specific aspects of LHCb operation. The single time sample data are also used to accurately determine the so-called sticky charge effect. This effect, where a fraction of the charge of a hit is carried over with a minus sign to the next read-out, is large for the channel numbers below 40 and is absent for channels with channel numbers above 80.

The analysis of efficiency and spill-over versus threshold shows that the Beetle can operate at about 97% efficiency with 10% spill-over hits. The Beetle pulse shape and the threshold settings need to be tuned together to get an optimal performance for LHCb.

Chapter 5

The LHC computing challenge

The data flow generated by *LHC*b and the other experiments is estimated at about 4 Peta byte per second. After the online selection, the data flow is reduced to about 40 Peta byte per year. Over their lifetime the LHC experiments will produce the largest dataset ever, in the order of Exa bytes in the year 2020. Since the report of the 'Hoffman review' to the LHCC early 2001 [68], the handling of that data is generally considered one of the most serious operational challenges LHC is facing.

At the moment an enormous amount of effort and resources is being invested in the development of large-scale distributed computing environments that can cope with this challenge. The computing needs exceed the resources available at any one of the LHC constituent research labs and universities. However, all of them put together should provide sufficient hardware in the not so distant future. As a result, one would like to combine this hardware into a giant pool of resources, which of course introduces organizational problems. Data and applications need to be distributed efficiently and the aggregate system must be simple enough, such that resources do not sit idle while humans struggle to configure and debug computers. These large-scale distributed computing environments that cross administrative boundaries are also known as computational and data grids, depending on their primary objective.

Already now a large amount of computation is needed to be able to understand the response of the LHC detectors. The physics interactions of the colliding beams need to be mimicked by the simulation of random particle collisions. These events are fed to a simulation of the detector geometry, which results in a trace of all of its output signals. These traces are fed into a reconstructor that infers the nature of the original collision events. The first phase of this computing activity is known as Monte Carlo production. Billions of simulated events are needed to analyze the signal and background of the various decay channels.

Experiments prefer to generate these events in large scale productions, called data challenges. A large number of resources can temporarily be dedicated to such a data challenge when it is run on a distributed system. These systems therefore provide the means to produce a large number of events on a short time scale. Monte Carlo production is normally run by a dedicated group within each experiment, which in general has a reasonable amount of control over the resources it uses. Hence, the organizational structure of Monte Carlo production is not nearly as complex as that for analysis, where each individual physicist is likely to be interested in access to a large amount of data and computational resources. However, Monte

Carlo production is the perfect test case for the more complex situation of data analysis.

Here, three problems are addressed that have so far limited the adoption of grid computing to highly specialized experts. First, users are used to batch systems, and the current prototype interfaces to grid software are often more complex and different from those in batch systems. Second, users are used to having transparent file access, which grid software does not conveniently provide. Third, efforts to achieve wide spread coordination of computers while solving the first two problems is hampered when clusters are on private networks. The proposed solutions for these problems are based on experiences with two applications that run on distributed systems.

The first application, the *LHC*b Monte Carlo production system, was designed from scratch to be deployed in a distributed environment and includes features that address the problems mentioned here. First, the complications of managing a distributed system that crosses various administrative domains are hidden behind a web interface that provides an easy way to access and use the application. Second, the issue of transparent file access is solved by the automatic transfer of the appropriate files to the distributed resources. Third, the restrictions caused by private networks can be circumvented if the nodes in these clusters are allowed to establish outbound network connections. The clients on the machines in the private networks only need to retrieve information from machines in the public network, which means connectivity is not a problem as long as these clients initiate the communication channels. Section 5.1 starts with a description of the features of this application, followed by a discussion of its deployment. First a traditional approach is discussed, where site managers are responsible for the initial deployment of the application and integration with their local batch system. Second a grid approach is discussed, where the application is distributed via the infrastructure of the European Data Grid (EDG).

A more general implementation of the proposed solutions can facilitate the adoption of grid computing by a broader public. This is the main motivation behind the way the second application is deployed as discussed in section 5.2. First, a user interface to access distributed resources is constructed around a standard batch system, which provides a familiar interface for the user and hides the complexities of dealing with various administrative domains. Second, transparent access to files is provided by a so-called interposition agent. This extra layer between the application and the operating system handles all file I/O of the user application and resolves the complications caused by distributed filesystems. Third, in contrast to the *LHC*b Monte Carlo production system, the batch interface does need inbound access to the clients on the machines in the private networks. This is achieved by automatically rearranging the way connections are established. Any attempt to connect to a node on the private network is reversed by a connection broker. In this way the communication channels are always initiated by the nodes on the private networks and the peer-to-peer application can operate as if symmetry in internet connectivity is recovered.

The power of this configuration is demonstrated in section 5.2.4 with the deployment of the complex BaBar [25] Monte Carlo production system on a computational grid, without any modification to the application itself.

5.1 Monte Carlo production

The *LHC*b Monte Carlo production system is named DIRAC, 'Distributed Infrastructure with Remote Agent Control' [69]. It has a client/server architecture based on a computing infrastructure distributed among the collaborating institutes. The communication across a network for accessing and updating the databases is accommodated via the *monitoring* and *cataloging services*. The databases contain information for production management, bookkeeping (the metadata catalog) and software configuration. Each production site installed a *site agent* to monitor the site-local batch queue, interrogate the production database for any outstanding production requests and to submit jobs to the batch system of the production site. The site agent checks and, if necessary, installs any software that the job requires automatically. After each job has processed the events, the site agent transfers the output data and updates the metadata catalog. DIRAC has been successfully installed at 18 collaborating institutes.

5.1.1 Functionality

The Monte Carlo production system provides the following functionalities.

- Definition of production tasks.
- Software installation on production sites.
- Job scheduling and monitoring.
- Data bookkeeping and storage management.

The last three of these operations are automated in order to minimize the interventions by local production managers to maintain the system. This is an important feature for the *LHC*b collaboration, considering the limited amount of dedicated manpower.

The job scheduling mechanism of DIRAC can be characterized as a "pull" approach. This means that the computing resource is actively seeking tasks to be executed. Once the resource becomes available it sends a request to *the production service*, a management process run on a central server. This service chooses a job according to the resource capabilities and then serves it in response to the request.

There are advantages and disadvantages of this method in comparison to the "push" approach, where the production service actively maintains the status of all computing resources and schedules the jobs accordingly. The most important disadvantage is the lack of information available for the production service. The status of a computing resource is not available for the job scheduler. As a consequence, an individual choice to combine a job with a resource might be suboptimal. The "push" method uses the information of all resources, but to schedule the optimal combination of resources and jobs it needs to investigate all resources for each individual job. As a result, the number of matching operations is equal to the number of available jobs times the number of resources, which can result in scalability problems. The "pull" method just matches the first suitable job with the first available resource and vice versa. Hence, for this Monte Carlo production system, where all resources are capable of running the jobs, the number of matching operations is equal to the number of jobs. Due to this straightforward job scheduling approach it becomes easier to fully exploit the available distributed computing resources at all the participating sites. Additional advantages of this

approach include a natural way of load balancing, since the more powerful resource will simply request jobs more frequently[1]. Furthermore, administrative overhead is reduced, *i.e.* new production sites can be incorporated without much effort, since the requests send to the *production service* are similar for all production sites and little or no extra information is needed to serve those requests.

5.1.2 Architecture

The DIRAC architecture is presented in Fig. 5.1. It can be divided into 4 geographically and logically distinct parts.

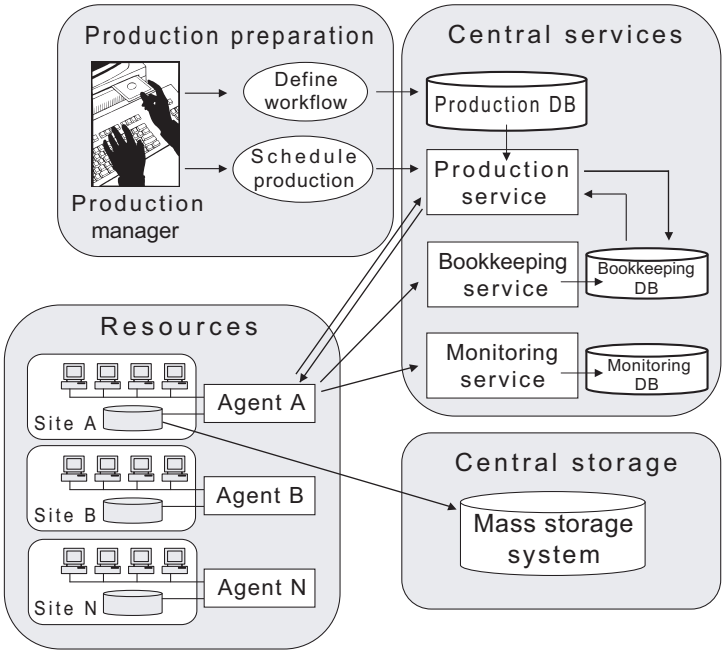


Figure 5.1: The DIRAC architecture. The arrows indicate the job information flow. Tasks are in ellipses, processes are in boxes and groups in boxes with rounded corners.

The production preparation is the part where human interaction is required. A (human) *production manager* defines a workflow describing the sequence of applications to be executed together with all the necessary application parameters. These include software version, application options, input and output data types, etc. These definitions are stored in the production database. The production manager makes sure that the required software is available in a central release area (in this case at CERN) upon requests of the production sites. Now, the *production manager* gives an assignment to the *production service* to produce the requested number of events in a production run. The *production service* checks the compatibility of the sequence of applications specified in the workflow and verifies the application parameters before it puts the jobs of the production run into a queue, to wait for available resources.

[1] In fact, the site agents on the production sites request new tasks from the production service when the occupancy of their batch queue drops below a given level, thus making sure to consume the maximum available amount of CPU time for that site.

The central services and the production database form the core of the production system. The *production database* contains all the information describing the production tasks as well as the job status parameters. It is implemented as an Oracle database running on a CERN server. The site agents communicate with the *production database* via an interface that is part of the *production service*. This service receives requests for jobs, checks the resource capabilities and serves the site agent with an available job. During the job execution the site agents will contact the *monitoring service* to report on the job progress. The *monitoring service* stores these messages in the *monitoring database* and renders them on a dedicated web page. After the job is completed, the site agents will contact the *bookkeeping service* to register the newly produced datasets. The metadata information is passed to the *LHC*b *bookkeeping database*, either automatically or by the production manager.

The site agents implement the previously discussed "pull" job scheduling paradigm. They are responsible for the complete job life cycle at the remote site. When the occupancy of a local batch queue drops below a given level, the site agent requests the *production service* for a new task. The site agent checks if the required software versions for this job are installed at the remote site and, if they are not, automatically fetches them from the release area and installs them. The site agent then submits the job to the site-local batch system of the production site and monitors its status. The job status is reported to the *monitoring service*. The job executes and writes its output to the storage system of the production site. When the job is completed, the site agent ensures that the output datasets are transferred to the *central storage*[2] and notifies the *bookkeeping service*. The bookkeeping service passes the storage information to the *bookkeeping database* when the transfer is completed successfully.

The communication between the services and site agents is implemented using the XML-RPC protocol. This protocol is a reliable standard for network communication and allows for efficient software development since it has an implementation in *e.g.* the Python language. The site agents can run as continuous user processes or as periodic jobs. They require an outbound network connection for communication with the central services and storage. Special attention was paid to the transfer failures of large datasets to the central storage, since it was known from previous experiences that this is the most delicate part of the job life cycle and required continuous baby sitting and manual interventions on a daily basis. DIRAC has an automatic recovery mechanism: in case of transfer failures, the data stay cached on the production site. Another transfer attempt is made during the next invocation of the site agent. This cycle continues until the transfer is successful.

5.1.3 Performance

The first *LHC*b Physics Data Challenge took place in February-April 2003. The goal of this production run was to provide sufficient Monte Carlo data to evaluate the *LHC*b physics performance in the Technical Design Report. This Data Challenge was the first time the DIRAC system was used in full scale production. In two months of continuous production 36,600 jobs were executed, each successfully completed job ran between 20 to 40 hours

[2] In principal mass storage does not need to be centralized, using newly developed tools for data grids [70]. However, since the *LHC*b community at the moment still uses CERN machines with classic file I/O interfaces (RFIO/FTP/AFS/etc.) to access the Monte Carlo data, a central storage is a technical necessity to make this data availability to them.

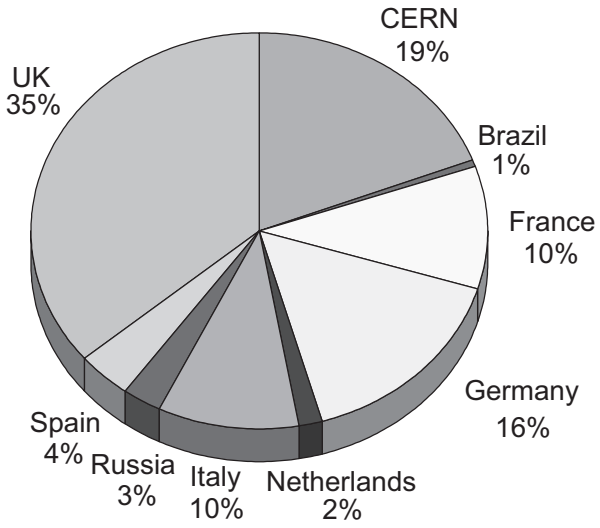


Figure 5.2: CPU time contribution of the LHC*b production sites to the Data Challenge 2003.*

depending on the CPU type of the host machine. The success rate was 92% (34,000 jobs) and almost 55 million events were produced. The failures were due to either *LHC*b software errors (2%) or problems at the production sites (6%). In total about 20 TB of data were stored in the central database, divided over 250,000 files. During several weeks about 1,000 jobs ran simultaneously on 18 production sites in 9 different countries. The CPU time contribution of the various institutes is shown in Fig. 5.2, specified by country.

During the production there were several *LHC*b Monte Carlo software updates, which were automatically installed at each site. The system proved to be stable and efficient. The goals of the Data Challenge were met ahead of schedule. These goals were conservatively based on our previous experiences with distributed production.

5.1.4 Expanding the production via a computational GRID

Over the next few years, the High Energy Physics, Earth Observation and Bio-Informatics research communities need to access and process large quantities of data that exceed the present quantities by at least an order of magnitude. Present systems will encounter performance problems or will be unaffordable for the participating universities and institutes when they are scaled to support the required amounts of users, data, processing and networking.

At the moment large general purpose resource facilities are under construction that can fulfill the requirements of the future applications of these communities. The most important advantage of these facilities is that resources can be shared. Each experiment has large fluctuations in their resource usage, so the duty cycle of their resources can be much higher when they serve multiple experiments. In business models this principle is known as ‘peak shaving’, which is one of the best ways to increase efficiency.

Although the combined resources of all these facilities will provide sufficient hardware in the not so distant future, they are located at geographically distinct locations and part of different administrative domains. As a direct result, users will be relatively anonymous. The combination of valuable resources with anonymous users is of course an interesting target

for malicious actions, which makes access control one of the primary concerns.

The European DataGrid (EDG)[3] project (started January 2001) addressed the challenge to manage these resources. The main objective of this test facility was to provide software and services to enable access for physicists to very large volumes of archived data and to a large number of shared heterogeneous computational resources, widely distributed across different organizations and centers over Europe and the rest of the world.

Already now, a large number of resources is available via the tools provided by EDG. The Monte Carlo production capacity drastically increases when the resources offered by these facilities can be exploited. As a proof of concept, DIRAC was tested on the EDG testbed. The *site agent* encapsulates all the functionality required to steer the execution of jobs on a remote site, which makes DIRAC well suited to be deployed on a computational grid. As far as the interface to the central services and databases is concerned, a computational grid is just another remote production center.

First some terminology is discussed of distributed computing in general and the EDG project in particular.

Public Key Infrastructure (PKI) Asymmetric cryptography which allows users to communicate securely without the need to exchange an encryption key in advance. This infrastructure forms the basis of the grid security infrastructure (GSI).

Key pair (public and private) Keys based on the public key infrastructure (PKI). The private key is kept secret by the user and the public key is available to the world. Data encrypted with the public key can only be deciphered with the private key and vice versa.

X.509 certificate A statement possessed by an entity (hardware, user, etc.), digitally signed by a certification authority, which declares the specific value of its public key for use in public key infrastructures.

Certification authority A trusted third party that digitally signs the X.509 certificates to confirm that the identifier contained in the certificate is a truthful representation of the identity that possesses the associated private key.

Testbed A distributed system that allows for testing of the grid software and resources by the high energy physics community. The experiments deploy their applications on these sites with the objectives to test the operability of the grid middleware[4].

EDG user interface Machine that runs the user interface client tools, which allows the user to interact with the EDG testbed. This is typically a machine the user logs into to submit jobs to the grid and to retrieve the output from those jobs.

Input sandbox A set of files that needs to be transfered automatically from the user interface machine to the grid resources by workload management tools, as part of the job submission (*i.e.* before the job can start).

[3] See http://www.eu-datagrid.org.

[4] Toward the end of the EDG project in March 2004, the testbeds reached production level quality and were used by the experiments to obtain sufficient resources to produce the high quality results they require from their applications. The definition of the term *testbed* varies widely across projects.

Worker node These are the nodes on which user computations are actually performed. They are typically managed via a local batch system, but these details are handled by grid software, relieving the user of having to keep track of site details.

Job description language (jdl) All information that is necessary to schedule and execute a job is specified with the job description language. This includes a specification of the files needed for execution (*i.e.* the input sandbox), environment variables and required resources.

Resource broker Machine that accepts user jobs from the user interface. It matches the jobs requirements to the available testbed resources and dispatches the jobs.

GSIFTP Standard file transfer protocol (FTP), enhanced to use the grid security infrastructure (GSI) service. It optionally provides high-performance features, such as parallel data transfer.

Storage element Node that provides uniform, high-level access to data storage. The storage element may control large disk arrays, mass storage systems and the like; however, the interface shields the user from the differences between these systems allowing uniform access.

Data set A data set can be any sort of collection of information, *e.g.* histograms, (sets of) files, event collections, *etc*. Once a data set is registered on the grid, it lives forever unless explicitly deleted.

Replica catalog The replica catalog contains information about the location(s) of data sets. It provides a consistent and up-to-date mapping between symbolic names and one or more copies on the physical storage systems.

Replica manager Currently, coordinates the interactions with the storage resources and replica catalogues while moving data. The ultimate goal is to optimize data access, which can be achieved via data replication (*i.e.* identical copies of data are generated and stored at various sites). This can significantly reduce data access latencies. The replica manager would then be responsible for replicating data and updating the replica catalog.

Figure 5.3 shows the deployment of DIRAC on the EDG *testbed*. A *submission agent* runs on a machine with an *EDG user interface* installed and acts as a gateway between the DIRAC production system and the EDG computing resource pool. The submission agent requests a task from the production service and packs it in an *input sandbox* together with a script to be executed on the *worker node*. Furthermore, the name of this script that needs to be started by the *worker node* is specified in the *job description language* (jdl), which is understood by the EDG services. The job is then submitted to one of the EDG *resource brokers*. The *resource broker* interacts with the authentication/authorization service of EDG to perform user authentication and authorization[5]. Once the job is authorized to execute on a *worker node* it installs a site agent, which effectively turns the *worker node* in a *LHC*b

[5]The authentication service of EDG allows users, processes and machines to prove their identities. This includes the support of single sign-on, whereby users can authenticate once and be authorized to

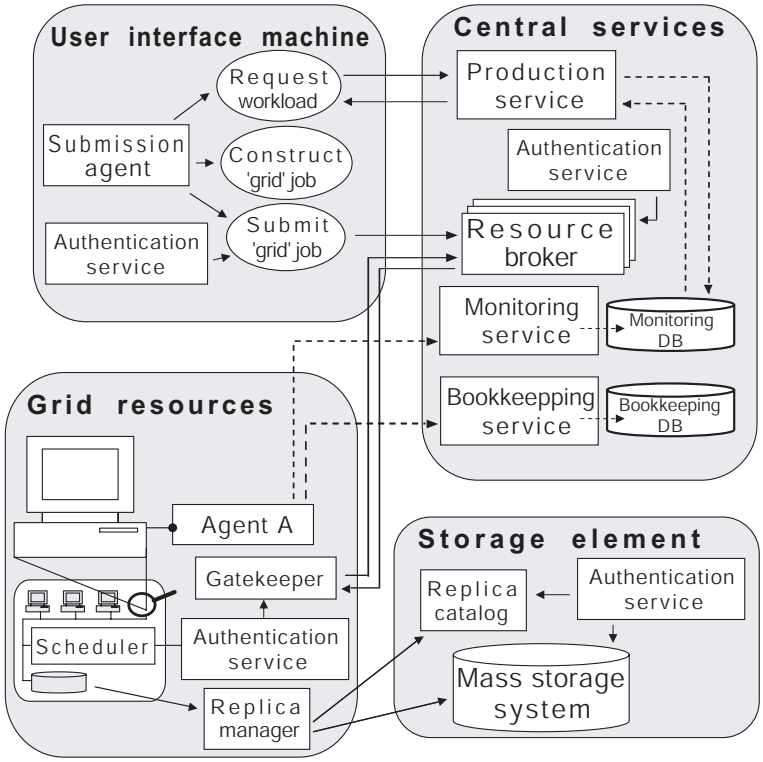


Figure 5.3: Deployment of DIRAC on the EDG testbed. The solid lines represent the flow of information associated to the operation on a computational grid and the dashed lines show the remaining parts of the traditional job information flow as shown in Fig 5.1. Tasks are in ellipses, processes are in boxes and groups in boxes with rounded corners.

production site. The production job is started and from this point onward all the operations necessary to steer the job execution are the same as for any other *LHC*b production site[6]. The site agent uses the *replica manager* to transfer the files to a specific *storage element* with *GSIFTP*. The selected *storage element* runs an extra process such that each file that arrives is immediately copied to the central storage of *LHC*b at CERN. The bookkeeping information about the files is provided to both the EDG *replica catalog* and the *LHC*b bookkeeping database, achieving easy access to the data for users operating at CERN or via the EDG services.

The success rate of the jobs submitted to EDG was between 25 and 60%, where the job length was correlated with the probability that it would fail. Failures were primarily caused by instabilities in the prototype grid software. Nevertheless, it was demonstrated that it is straightforward to run the *LHC*b Monte Carlo production software in a grid environment.

access many grid resources. The secure environment for this authentication is based on a *Public Key Infrastructure* (PKI). Each user and each service owns a credential, consisting of a *pair of keys* (one private and one public), to prove their identity. The user or service identity and their public key are included in a *X.509 certificate*, signed by a trusted *Certification Authority* (CA), which guarantees the association between the public key and its owner.

[6]It is not really necessary to specify the workload in advance. The site agent could request its workload from the production service when it starts to run on the *worker node*. This would allow for a more opportunistic scheduling approach.

Even a tailored application like DIRAC encountered problems that limited the usability of the grid resources. One of them is especially interesting, since it puts a handle on more widespread usage of grid computing by ordinary applications. The experiences gained with DIRAC showed a great importance of the possibility for the running jobs to send messages to the central services. The existence of a communication channel makes it possible to dynamically customize the software environment on the worker nodes and to have fine-grained information of the job progress [69]. It requires the outbound IP connectivity of the working nodes, which is highly debatable because of security concerns. If this connectivity is not available, proxy infrastructure needs to be developed to successfully implement similar tasks (*e.g.* RGMA[7]). However, if outbound IP connectivity of the worker nodes does become a standard in grid computing, we can use it to transparently connect any complex application to distributed systems in a similar way as DIRAC does. A possible implementation is discussed in the next section.

5.2 Providing transparent access to grid resources

The adoption of grid technologies can significantly be accelerated when individual physicists are able to run their own analysis on grid resources. Their satisfaction with grid computing heavily depends on the amount of effort required to deploy an application on such a distributed system. This effort is drastically reduced when users can submit their jobs through a well-known batch interface, one that does not require extra knowledge to run an application on resources in different administrative domains. Access to grid resources should therefore be transparent, much like submitting a job to such a batch system.

Authorized users who want to run applications on EDG resources interact with the *resource broker*, shown in Fig. 5.3, to submit their jobs. This component is discussed in more detail in section 5.2.1. It is based on *Condor-G*, a tool that uses inter-domain resource management protocols to provide resource discovery and resources access across administrative domains [71]. As will be shown, the resource broker provides the users with only limited capabilities to manage grid resources and is therefore the main obstacle for the construction of a transparent user interface.

A solution to this problem is a technique called *gliding in*, which is discussed in the same section. This technique allows the remote resources to become an integral part of the user's batch pool. This 'virtual batch system', like most batch systems, needs bidirectional communication with its worker nodes. However, in contrast to normal batch nodes, grid resources can often not be accessed directly by the user's batch system. Therefore, connection issues are handled by a generic layer in between the batch system and its system calls, which makes it possible to use the batch system across firewalls and private networks. This *generic connection broker* is discussed in section 5.2.2.

Although gliding in provides transparent access to grid resources, many applications will not be able to complete successfully on these remote machines without heavy modification. Note that DIRAC was used successfully on a computational grid because it was developed from scratch, by computing specialists, to be deployed on a distributed system. Many physicists will not have the luxury of redeveloping their application from scratch, or they use

[7] See http://www.r-gma.org.

the grid infrequent enough to warrant a serious grid-specific development effort. Hence, it should be made possible to deploy an unmodified application, to address the cases where one benefits from grid computing only if efforts to deploy an application are minimal.

Let's elaborate on the reasons why most applications will fail to run on a distributed system. Installing most software on a new cluster is a labor-intensive process that is not suited for automation: executables, scripts, and libraries must be unpacked and installed; environment variables and other settings must be configured; database structures must be initialized; dependent software must be discovered and installed. Users are expected to have administrative privileges on the machines they use to run many sorts of production systems (this is a general problem, not specific to grid computing). Some software expects a uniform user database across multiple machines; this is an impossibility on a computational grid. Furthermore, the nature of a distributed environment ensures that network outages and performance variations are common events.

These problems will not fade away on a short time scale. Physicists typically write their applications on standalone machines, making liberal use of complex and powerful utility libraries and programming environments. By re-using existing tools, they are able to concentrate on their craft rather than reinventing computing from the ground up. Software is created, debugged, and validated on ordinary workstations long before any thought turns to distributed computing[8]. Still, the success rate of remotely executed applications is as least as important for user satisfaction as transparent access to the resources.

A solution for this problem is the introduction of an extra layer in between the user application and its system calls. In section 5.2.3 an interpositioning technique based on the debugger trap is discussed to provide the user with transparent file I/O over a wide area network without any modification to the user application or the underlying computing systems. The *interposition agent*, called Parrot, provides the user application a virtual environment with the illusion of a local system, independent of the actual location of the resources it uses.

The components mentioned in this section allow to use grid resources as if they were local while providing transparent access to files, even when private networks intervene. As a realistic test, the deployment of an application into a real distributed environment is described in section 5.2.4. This application, BaBar Monte Carlo production - SP5[9], is representative of the applications described earlier; it consists of multiple processes and complex libraries, poorly suited to most distributed computing and file systems. In particular, the problem of working through an aggressive firewall is addressed.

5.2.1 Resource management

One of the primary objectives of the grid user interface is to facilitate the interaction between the users and the resource broker. In order to demonstrate the functionality offered by this resource broker and to understand the limitations it has with respect to resource management, it is presented here in a bit more detail.

When an authorized user requests to run jobs on grid resources, the resource broker responds by creating a new grid manager process to submit and manage those jobs, as shown in Fig. 5.4. The grid manager is based on *Condor-G* and uses the protocols defined by the

[8] Although in software development this might be considered as bad programming practice.

[9] http://www.slac.stanford.edu/BFROOT/www/Computing/Offline/Production/userguide/userguide.html

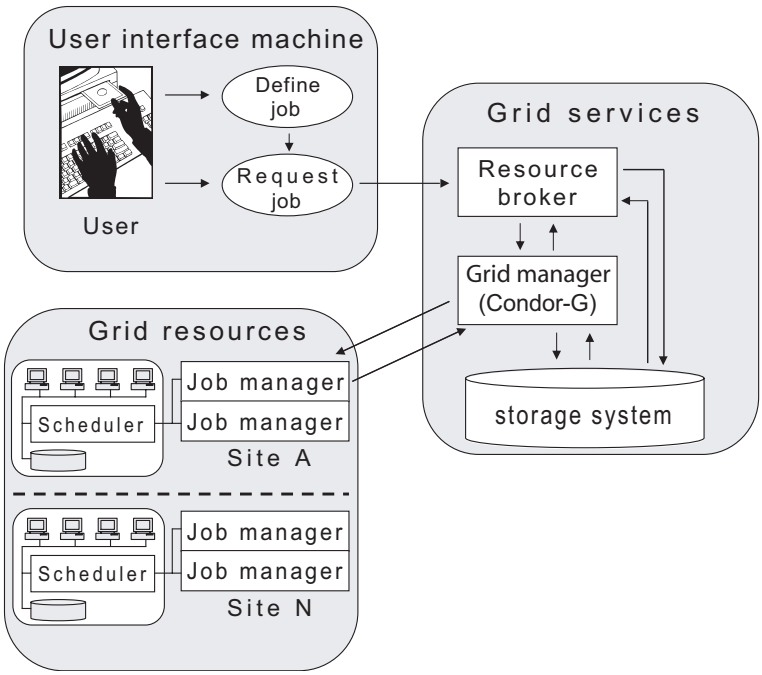


Figure 5.4: Condor-G functionality. Tasks are in ellipses, processes are in boxes and groups in boxes with rounded corners.

Globus Toolkit [72][10] to communicate with the grid sites. One grid manager handles all jobs for a single user and terminates once they are complete. All authentication requests are handled by the grid manager via the Grid Security Infrastructure (GSI), which allows for single sign-on based on a Public Key Infrastructure (PKI). The submission of jobs to a remote site is communicated via the Grid Resource Allocation and Management (GRAM) protocol. On the grid sites each received job results in the creation of a job manager process[11]. This process handles the job life cycle at the grid site. It connects back to the grid manager, applying GSI mechanisms for authentication and retrieves the job's executable and input files using Global Access to Secondary Storage (GASS). The standard output and error of the jobs are streamed real-time via the same route. The job manager process submits the job to a scheduler running on the same grid site and keeps the grid manager up to date about the job status.

The techniques described here allow for the construction, submission and monitoring of jobs on remote resources. Condor-G provides fault tolerance that handles the complication of a distributed environment that spans multiple administrative domains. However, the system does not provide a transparent user interface due to two barriers:

- The users need to provide a lot of information to the grid manager for job submission and for the retrieval of job information. This information is required to deal with the complexity of the grid infrastructure. Nevertheless, by confronting the users with this complexity a barrier is formed between the users and the grid manager.

[10]The Globus Toolkit is a de facto standard for grid computing

[11]Note that this implementation causes scalability problems when a large number of jobs is submitted to the same site (and has been changed in the successor of EDG, the LHC Computing Grid – LCG).

- The second barrier is caused by the interaction between Condor-G and the various grid sites: jobs will end up in different queues, often with different types of scheduling policies and priority schemes. The user has only limited and indirect job control possibilities, due to the communication route: resource broker→grid manager/Condor-G→scheduler→job. On top of this, the diversity of the queues makes it difficult to accurately interpret information about *e.g.* the job status.

These two problems are actually symptoms of a more fundamental issue. Like most batch systems, grid resource management tools combine the stages of job submission and resource allocation. When the user submits a job, the resource broker starts looking for a resource that matches the request. However, since resource allocation is much more complicated on a grid infrastructure than it is on a batch pool, this results in a more complex user interface as well as in difficulties to interpret the available information. Instead of treating these symptoms, the solution proposed here is based on a method that disentangles the stages of job submission and resource allocation.

The first part of this separation is accomplished by using a batch system that manages the available resources for a particular user or user group. It provides a job queuing mechanism, a scheduling policy, a priority scheme, resource monitoring and resource management [73]. The users submit their jobs to this batch system, which places them into a queue. In the rest of this chapter, this batch system is called 'the user's batch system'. The user's batch system chooses when and where the jobs are run based upon its policy and monitors their progress. Ultimately the user is informed upon completion.

The resources managed by the user's batch system are acquired from grid sites with a technique called *gliding-in*, presented in Fig. 5.5. In the gliding-in scheme the jobs submitted to the resource broker contain requests to execute batch system processes instead of user jobs; thus providing the second part of the separation between job submission and resource allocation. When these processes are running on the worker nodes of the grid site(s), they will contact the batch resource manager and appear in the user's batch system. In this sense, gliding in provides a virtual batch system: the acquired worker nodes provide all the features of normal batch nodes. Users can submit their jobs to these resources directly via their batch system, without the use of grid mechanisms. The user's batch system is responsible for the management of the worker nodes until the moment that the claim on the grid resources expires. The idle jobs in the user's batch queue are awaiting either the allocation of new resources or for one of the already allocated resources to become available.

Still, in the resource allocation stage one inevitably has to deal with the grid user interface. In Fig. 5.5 this part is contained in the *resource acquirer*, which interacts with the resource broker. Although it seems like the problem just shifted, only limited interaction is required between the user and the resource acquirer, which might even be implemented as a completely autonomous process. The resource acquirer can guarantee optimal queuing times to its users by submitting batch processes to all remote resources, cancelling the ones it does not use. This would prevent a job from waiting at one remote resource while another resource capable of serving the job is available[12]. Batch processes on remote sites

[12]This can also be accomplished by directly submitting the user job to multiple sites. However, the separation of the allocation and execution stage allows for a more opportunistic scheduling approach. It is not necessary to have the user job available when a request for resource allocation is submitted.

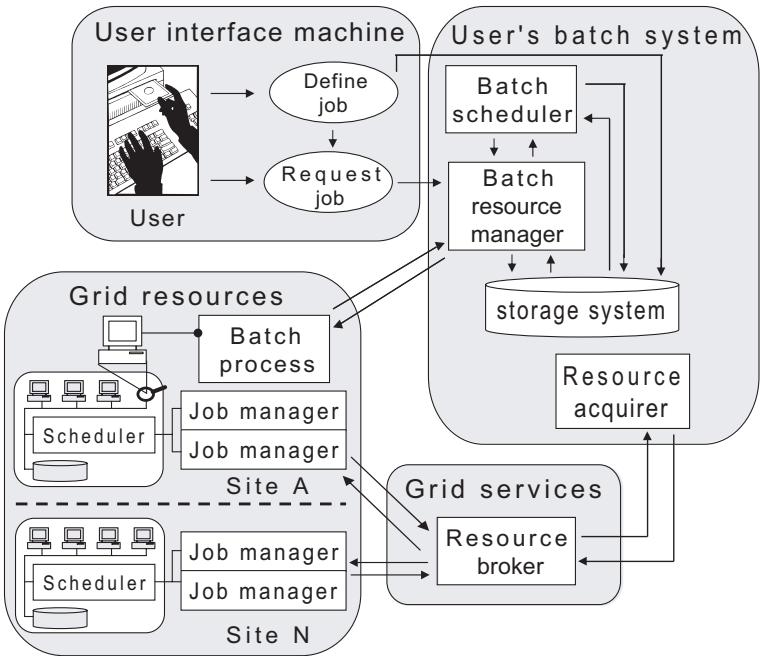


Figure 5.5: Gliding-in functionality. Tasks are in ellipses, processes are in boxes and groups in boxes with rounded corners.

shut down gracefully when they do not receive any jobs to execute after a (configurable) amount of time, thus guarding against runaway processes. The development of an intelligent resource acquirer that is able to collect information from the user's batch system and from grid resource planning and scheduling tools, would allow to optimize the acquisition process.

5.2.2 Dealing with network complexity

The effort to achieve the wide-spread coordination of computers discussed in the previous section is hampered when clusters are on private networks or behind firewalls. Nevertheless, private networks are common practice in grid infrastructures, since they provide easy network management and address planning as well as a solution to the IPv4 address shortage problem. Firewalls play important roles in protecting networks and are ready to play an even more important role as the security headquarters of integrated security systems.

The information required to start the batch process on the worker nodes of the grid sites flows without problems from resource acquirer to resource broker to job manager by means of grid mechanisms. However, when the batch process is started, it has to be managed from the user's batch resource manager. Batch systems in general assume the availability of bidirectional communication with the worker nodes. Although the worker nodes on grid sites usually have outbound IP connectivity, inbound connectivity is much less common. This prevents the user's batch resource manager from *e.g.* submitting user jobs to the worker nodes.

For the user's batch system to work across private networks and over firewalls the symmetry in the peer-to-peer connections needs to be recovered. The solution should not require any special privileges on the side of the grid resources. *Generic Connection Brokering* (GCB) [74] possesses these properties. An example of a system with GCB is shown in Fig. 5.6.

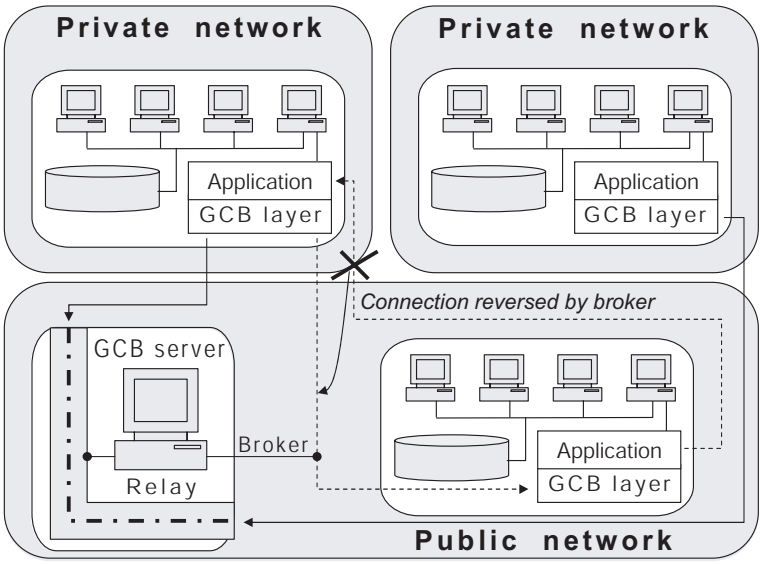


Figure 5.6: GCB functionality. Tasks are in ellipses, processes are in boxes and groups in boxes with rounded corners. The arrows and lines are explained in the text.

The system shown consists of one node on the public network and two on private networks. The applications are GCB enabled on both the public and private nodes, *i.e.* they are able to use GCB functionality to solve communication issues. Outbound IP connectivity of the nodes on the private networks is assumed[13].

The nodes in the private network send a register request to a GCB server that resides on the public network (solid lines). The server creates proxy sockets of the same type as the client sockets, binds them, makes them passive (*i.e.* they are configured to listen to incoming connections) and returns the addresses to the private nodes. From now on, the private nodes use these addresses as their network identity (*i.e.* whenever they need to inform other processes of their address, they send the proxy address instead of their real address). With respect to communication, there are two interesting situations that can occur:

1. The application running on the public node wants to connect to an application running on a private node (dashed arrow). However, the network identity of the application on the private node has been changed to a proxy address on the GCB server. So, without being aware of it, the application actually contacts the GCB server. The server decides, based on the network situation of both nodes, to reverse the connection (indicated with the cross, the curved arrow and the dashed arrow). The server passes this information to the GCB layer underneath the application on the private network. This GCB layer now initiates the connection, which is received by the GCB layer underneath the application on the public node. The communication channel between the two GCB layers forms a tunnel, that enables the applications to make the originally requested connection.

[13]This statement is not entirely correct. The nodes on the private networks need to be able to contact the GCB server and the GCB server needs to be accessible via a public internet address. If the GCB server manages only one private network, it could be installed on the head node and the private nodes would not need outbound IP connectivity. This construction would however require the ability to start a long-running process on the head node.

2. An application running on a private node wants to connect to an application running on a private node of a different private network. Again, the application contacts the GCB server, but now either cannot connect to the other. Therefore, GCB arranges that both parties connect to the server and the packets are relayed between them (dash dotted line).

The GCB server is implemented as a process that can run with standard user privileges. Since the server does not assume the possibility of initiating connections to the clients, it can be placed anywhere on the public internet. It maintains accurate information about the status of the clients using the heartbeat messages they send periodically.

When Gliding-in and GCB are combined, they are capable of extending the functionality of the user's batch system into private networks of other domains without the need of administrative privileges on any site. To combine them, GCB is implemented as a layer between the batch system software and its system calls. Since the GCB layer provides the same interfaces and semantics as the standard socket calls, the batch system software can be linked with GCB without modifications[14].

The GCB layer checks every incoming message from the network and decides to either pass it to the batch system, or handle it appropriately if it is a GCB command. GCB can use both the TCP and UDP protocol for communication. In our experience the end to end reliability of TCP makes this protocol preferable above the UDP protocol in a grid environment. Although the UDP protocol is in principle superior in performance, it silently fails to pass the stringent requirements of active components in some private networks (like network address translators).

GCB recovers connectivity which was disabled because of (in some cases) security reasons. Without intervention of network administrators, it may be deployed to any grid site that allows only outbound connections from the worker nodes. The GCB server is the weakest point in the system described. It maintains connections to all the acquired machines in the private networks, and represents both a single point of failure and a security concern. This stretches the security policy of many sites, maybe even to an extent that allows for abuse. However, distributed computing exists by the virtue of communication, hence hampering connectivity is not the way security should be provided.

Developments are on their way to provide GCB with a strong security mechanism. Furthermore, since the resources managed by the GCB server can span multiple private networks, an extended addressing scheme is required to uniquely identify all machines. This functionality is implemented in a new version of GCB, which will avoid the private IP address collisions we experienced [75].

A virtual batch system for the submission of user jobs was successfully created and extended across various administrative domains, private networks and firewalls with the techniques discussed in this section. The results are discussed in section 5.2.4.

5.2.3 Remote execution with interposition agents

Although the virtual batch system allows for easy resource management, it does not provide transparent access to files. For this an *interposition agent* can be used, which allows

[14]Note that GCB is *only* linked to the batch system software, which does the resource management. The user application will run exactly as it is provided.

modification of the file I/O of any application, without modifying the application itself.

An interposition agent is a piece of software that inserts itself between two existing layers of software in order to modify their discourse. By inserting an interposition agent rather than modifying an existing piece of software, we may measure, debug, and enhance an application without requiring intimate knowledge of its innards. An interposition agent has many uses in a distributed system:

- Integration of distributed resources. The most common use of an interposition agent is to connect an application to a new resource, such as a storage device, without requiring any special changes or coding in the application. For example, an interposition agent can allow an application to connect to a remote storage server as if it is an ordinary file system.

- Improved reliability. In general, remote data services are far less reliable than local filesystems. Remote services are prone to failed networks, power outages, expired (grid) credentials, and many other problems. An interposition agent can attach an application to a service with improved reliability. For example, it can emulate a reliable TCP connection across network outages and address changes or add reliability at the remote file system layer by detecting and repairing failed I/O connections. When combined with generic connection brokering, it can even provide bidirectional communication for peer-to-peer applications across private networks and firewalls.

- Private namespaces. Batch applications are frequently hardwired to use certain file names for configuration files, data libraries, and even ordinary inputs and outputs. An interposition agent can be used to create a private namespace for each instance of an application, thus allowing many to run simultaneously while keeping their I/O activities separate. For example, several instances of an application hardwired to write to ”output.txt” may be redirected to write to ”output.n.txt”, where n is the instance number.

- Remote dynamic loading. Although dynamic linking offers many technical advantages for programs that share code or data, it presents a number of practical problems. It is all too easy to migrate an application only to discover that needed libraries are missing, or worse yet, that the available libraries are the wrong version. An interposition agent can solve these problems by allowing an application to load libraries from a single, well-known server.

- Provide profiling and debugging information. The vast majority of applications is designed and tested on standalone machines. A number of surprises occur when such applications are moved into a distributed system. Both the absolute and relative cost of I/O operations change, and techniques that were once acceptable (such as linear search) may become disastrously inefficient. By attaching an interposition agent to an application, a user may easily generate a trace or summary of I/O behavior and observe precisely what the application does.

Parrot [76] is an interposition agent that provides the features discussed above for standard Unix applications. It observes and potentially modifies the interaction between an unmodified process and the operating system kernel using the standard Linux ptrace debugging

interface, which traps all system calls of the process. When used in a distributed system, Parrot provides the illusion of a user's home environment, including files, user identities and more. It can customize an application's environment to create a synthetic namespace formed from multiple remote services. In addition, it is able to deal with network outages, server crashes, and other failures that are endemic to distributed systems, thereby minimizing the effects on the application.

Although the notion of interposition agents is not new, they have seen relatively little use in production systems. This is due to a variety of technical and semantical difficulties that arise in connecting real systems together. For example, many different I/O protocols may be attached to an application, but few provide the full range of POSIX semantics expected by many applications. For this reason a dedicated protocol was created, Chirp, which provides the precise POSIX semantics that applications expect. Each Chirp operation is a remote procedure call from a client to a server. A Chirp operation is initiated by a client, which sends a formatted request. The server acts upon the request and sends a response. It is assumed that Chirp is carried over a stream protocol such as TCP. Authentication and authorization can be done through a variety of methods. Most interesting for the current setup is the GSI authentication as mentioned before, which allows for integration of the Chirp server with the batch system using the existing security infrastructure.

Parrot is an extension to an existing operating system; it augments file-handling capabilities without affecting the ability of a process to interact with other processes on the same machine or over a network. Parrot is considerably simpler than other tools like virtual operating systems such as User Mode Linux and virtual machines such as VMWare, both which require the user to build and maintain large filesystem images and all the elements of an isolated operating system in miniature. Parrot consists of a single executable measuring only 8.4 MB with all options enabled, and as small as 1 MB in minimal configuration.

Figure 5.7 shows the control flow necessary to trap a system call through the ptrace interface. Parrot registers its interest in an application process (*i.e.* the user application) with the operating system kernel. At each attempt by the application to invoke a system call, the host kernel notifies Parrot. Parrot may then modify the application's address space or registers, including the system call and its arguments (*i.e.* change local file I/O to a Chirp remote procedure call). Once satisfied, Parrot instructs the host kernel to resume the system call. At completion, Parrot is given another opportunity to make changes before passing control back to the kernel and the application.

5.2.4 An example: BaBar Monte Carlo production

The Monte Carlo production system of the BaBar high-energy physics experiment in progress at the Stanford Linear Accelerator Center is called SP5. Although the exact details of this application do not matter, it is interesting to know how SP5 operates at an abstract level. First, it loads the data that describes the configuration of the detector and the physics of particle generation. Once loaded, it enters a compute-intensive phase where it generates an arbitrary number of events that can each be summarized in 10-100 kilobytes. This means, in theory, SP5 has the right structure for distributed computing. The initial data can simply be distributed to a number of machines, production can be performed in parallel, and the produced events can be returned to a central site. Once initialized, any processor can produce

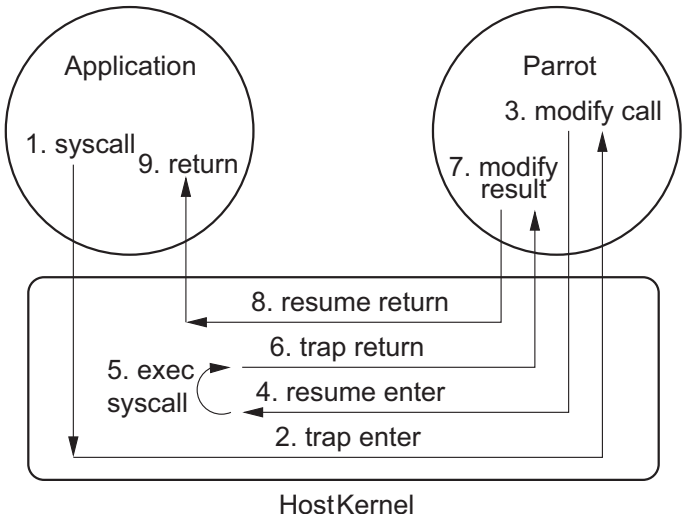


Figure 5.7: Interpositioning via the debugger interface.

an arbitrary number of events, so the number of processors can be chosen to balance startup time against desired throughput.

In practice, SP5 has a number of complexities that make it difficult to deploy in a distributed system. It consists of a number of executables and scripts, several dynamic libraries, the input configuration, and the output events. The production of Monte Carlo events is started with a 'wrapper script' that establishes environment settings, verifies the integrity of the files and invokes the program. It also makes use of several dynamically-loaded libraries, particularly the Objectivity database, which manages the configuration and event data.

Objectivity is a decentralized, cooperative database built on top of a filesystem. Consistency management, access control, and crash recovery are performed cooperatively by clients rather than enforced by a server. A minimal central server assists only with a locking protocol. To read the configuration data or write events, the client library requests a lock from the lock server, manipulates the files directly, and then releases the lock.

This structure is quite reasonable for a local area network, where network connectivity is more or less guaranteed, but is difficult to adapt to a truly distributed system. Moreover, the filesystem activity of the Objectivity client library cannot be carried over a standard distributed filesystem. The semantics of NFS clients are too weak for database structures[15], while the strict open-close semantics of AFS would result in data loss on the append-only transaction log[16]. Objectivity does have the capability to speak NFS directly to a server, bypassing the buffer cache, but deploying this requires superuser privileges at both the client and the server; an unlikely capability for the clients in a grid computing environment.

[15]Most of the kernel-level NFS clients have vague semantics. Data written to an NFS-mounted filesystem may stay in the buffer of the local machine anywhere from 3 seconds to several minutes, before actually being written to the server. This can result in database inconsistencies when multiple clients or processes are accessing the same files in the database.

[16]AFS only transfers complete files. A database like Objectivity often needs to open the transaction log, write to the end and close the file. Clearly, this results in AFS performance problems when the log is growing. On top of that, semantics problems occur when two processes attempt to append to the log file at once. Due to the Get-Put nature of AFS, the changes of only one of the processes will be accepted.

5.2.5 Deploying SP5 on a computational GRID

Figure 5.8 shows how the pieces of the distributed environment fit together. The configuration data and output events are stored in files managed by Objectivity on a specified central server. A central lock server process assists with mutual exclusion. A number of worker nodes are used to execute instances of SP5. Access to these worker nodes at various universities and institutes is obtained via the gliding-in technique, as discussed in section 5.2.1. No superuser access is required in these environments and only the standard grid software needs to be installed on the grid resources.

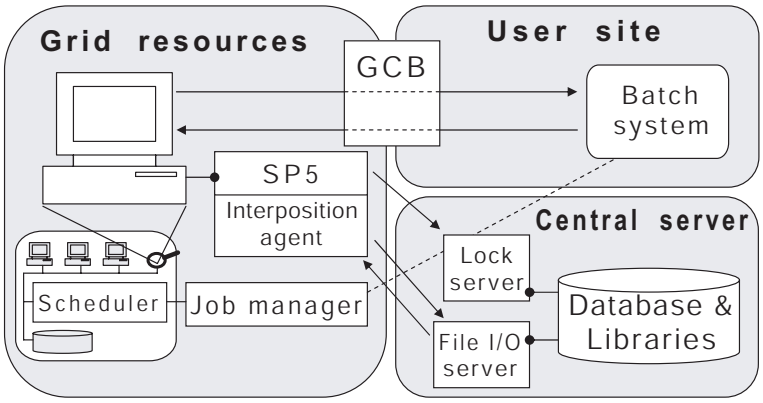


Figure 5.8: Deploying SP5 and Parrot on a distributed system. Tasks are in ellipses, processes are in boxes and groups in boxes with rounded corners.

Jobs are submitted to the worker nodes with the user's batch system. These jobs contain both the SP5 wrapper script and the interposition agent (Parrot). Parrot makes the database and libraries appear local for the application that runs on the worker node. This is comparable to an NFS client that mounts its root filesystem from a remote device. All executables, dynamic libraries, and other program components are loaded from the central server via the Chirp protocol. The SP5 filesystem operations are carried back to a Chirp file I/O server run with standard user privileges deployed at the central server. Parrot makes local copies of executables; this is a technical necessity, because Unix can only execute a program identified by a local file name. All data files are accessed remotely without caching, to avoid consistency problems in the database. Note that the application that runs on the worker node initiates all connections, so there is no need to link Parrot with GCB in this specific case.

In addition to the transfer of filesystem operations, a number of other small settings were necessary to fully emulate the home environment. For example, the Objectivity libraries examine the POSIX user identifier and host name in order to implement access control on the database. Because worker machines do not share a user database with the central server, Parrot is instructed to trap these system calls and change the results to match what would be seen at the central server.

Aggressive firewalls posed a serious problem to the deployment of this system. It is quite common for a computing cluster to be connected to the public Internet by way of a firewall and/or network address translator (NAT). In the clusters targeted by this application, the NAT permits cluster nodes to initiate outgoing TCP connections to the public Internet, but prohibits incoming connections. To translate external addresses into internal addresses,

the NAT must keep state about every TCP connections that it carries.

The problem arises because a NAT must eventually discard TCP connections that it perceives to be idle. Each connection consumes some state in the firewall, so it cannot keep them forever. The most aggressive NAT encountered discards TCP connections that have been idle for only one minute. When this happens, there is a double penalty: not only is the connection lost, but the NAT does not even return a message indicating that the connection was lost. The result is that both sides think the connection is present but unreliable, and retry up to their maximum timeouts, which can range from minutes to hours.

This problem was deadly to SP5. Once it initialized, the lock server connection was held open and idle, while the Chirp connection was only used for the output of each event, at intervals of slightly more than a minute. While SP5 was processing the first event, the NAT would discard the TCP connections. A short time later, the entire system would hang while attempting to write out the first event.

Although one would be tempted to discount this firewall as an aberrant device, reconfiguring its timeout cannot be considered a reasonable solution. For the same reasons software cannot be installed at the worker node and superuser privileges are not available, reconfiguring the network interior at will is not an option[17].

One solution to this problem is to change the network endpoints to generate enough traffic to keep the NAT state alive. For example, the networking stack at the central server can be modified to send TCP keepalives at the rate of several per minute. This technique was applied in order to preserve the connection between SP5 and the lock server. However, it is unsatisfying because it requires administrator privileges on at least one end and has a system wide effect, thus all sockets are affected.

A more comprehensive solution is to make the network protocol recoverable, so that the failure of the TCP connection becomes a harmless event. For example, as part of this project Parrot was modified so that a failed Chirp connection was recovered by reconnecting and reopening the needed files. With this recovery method, the Chirp connection was made *fail-fast*; hence, any delay of greater than thirty seconds was assumed to be a transient network failure and would result in disconnection and recovery. This solution is more robust than simply applying keepalives, *e.g.* it also tolerates the crash and recovery of the Chirp server. Note that the Chirp protocol had to be modified to accommodate for this recovery method, which is something that could only be done because this protocol is not wide spread and is used here for a specific implementation.

The Chirp recovery method circumvents a well-known problem in the design of distributed filesystems. Strict POSIX semantics require that an application holds references directly to files rather than names. That is, once an application opens a file by name, it keeps access to that file even if the name is deleted or renamed. Distributed file systems such as NFS and AFS solve this problem by exposing inode numbers to clients. When recovering from a disconnection, NFS and AFS clients can be assured of access to the correct files by referring to the inode numbers. Chirp cannot do this directly; the Chirp server is implemented on top of an ordinary file system and runs without superuser privileges, thus can only open files by name. However, the Chirp protocol can verify that the binding between names and inodes

[17] In fact, it was later discovered that these are the factory settings for the NAT in question. The idea of negotiating with a network administrator every time this model of NAT is encountered is not very tempting.

has not changed after a recovery by simply querying inode numbers with the *stat* operation. If they have not changed, then recovery is successful. Otherwise, recovery has failed, Parrot forces SP5 to fail immediately, and the batch system becomes responsible for re-starting the job from the beginning.

In the remaining part of this section the focus is shifted from the problem of recovery to performance issues of SP5. Table 5.1 shows the run-times of SP5, gradually increasing the logical (*i.e.* method and protocol) and physical distance between it and its data on the central server. As discussed earlier, SP5 begins with an I/O-intensive startup phase, and then settles into a CPU-intensive phase of configurable length. As the distance increases, the I/O-intensive phase pays an increasing price, but the CPU-intensive phase is less sensitive.

Table 5.1: Performance of SP5 and Parrot deployed in a distributed system

distance	*method*	*protocol*	*CPU*	*time to initialize*	*time per event*
local	os	files	1 GHz	446 ± 46 s	64 s
local	parrot	files	1 GHz	668 ± 26 s	65 s
local	parrot	chirp	1 GHz	777 ± 48 s	66 s
lan	parrot	nfs	1 GHz	4464 ± 172 s	113 s
lan	parrot	chirp	1 GHz	4505 ± 155 s	113 s
wan	parrot	chirp	2.5 GHz	6275 ± 330 s	88 s

The first line of Table 5.1 shows the performance of unmodified SP5, running on the same machine as where the data reside. The application is run in 'validation mode', producing additional histograms to cross check the results. Furthermore, it produces a full debugging output so that the correctness of the output can be verified. As a result the production is approximately a factor of 5 slower than the standard production on this machine. The average and standard deviation of initialization times are shown along with the average time to process an event. It initializes in 443 seconds and then processes one event every 64 seconds. Each measurement of the initialization time is the result of 10 trials. The time to process one event is an average of 2,000 events. A small numbers of outliers beyond 5σ were attributed to unrelated network traffic and discarded. Each successive row adds one component in order to measure its contribution. The second row adds Parrot, but without any remote I/O or other features; SP5 just accesses local files through Parrot. The third row adds Chirp, but without a network; SP5 accesses a Chirp server on the same machine using Parrot. As can be expected, both Parrot and Chirp slow down initialization, but have little effect on event processing.

The fourth and fifth rows show the performance of SP5 accessing its data over a local area network (latency $130 \pm 10\mu$s). In the fourth row, SP5 is using a kernel-level NFS client to access Objectivity's files, ignoring potential consistency problems due to caching. In the fifth, SP5 is using Parrot and Chirp to accomplish the same task safely without a cache. Although initialization is an order of magnitude slower than the unmodified case, the performance of Chirp is comparable to NFS. The overhead is more a function of the network than of Parrot or Chirp. Note that the 'time per event' increased with a factor of two. This is a known limitation of SP5 and is caused by writing the events into the central database. The new versions of the BaBar Monte Carlo software no longer write the events to a central database during production. Instead, the results are cached on the worker nodes and transferred

at the end of the job.

The final row shows the performance of the complete system as depicted in Figure 5.8. SP5 accesses its data over a wide-area network (latency $654 \pm 50\mu$s) via the firewall as discussed above. Notice that the performance numbers are not directly comparable, as the CPU is about 2.5 times as fast as the others. (The CPUs in a real-world distributed system are rarely identical.) However, the same qualitative result as the other lines may be seen: initialization is slow, but event processing is reasonable.

Overall, the BaBar experiment must process billions of events to complete the required simulations. In the worst case of accessing data over a wide area network, the cost of computing events equals the cost of initialization at only 70 events. Given that a typical instance of SP5 processes 10,000 events, the cost of remote execution, while significant, can be amortized across a large run.

The batch system provides the user with a familiar interface to distributed resources. The complexities that arise from resource management in a system spanning multiple administrative domains and private networks can effectively be hidden by Condor-G, the gliding-in technique and GCB. Interposition agents bridge the gap between applications and systems when neither are available for modification. By raising the level of abstraction on which an application executes in a batch system, a transparent and reliable environment is provided, even in an unreliable distributed system. Deploying a complex application into a distributed system is quite feasible for an ordinary user with these tools as shown with the BaBar Monte Carlo production system, SP5.

Chapter 6

$B_s \longrightarrow J/\psi\phi$ analysis

So far, various aspects of the *LHC*b experiment were discussed individually. Here, the combined performance of *LHC*b and the resulting accuracy of the measured physics parameters are presented. This chapter starts with a discussion of the applications responsible for the simulation and processing of data. The main design features of the software framework are presented, followed by a discussion of the various components which characterize the data handling in *LHC*b.

Next, the selection of $B_s \to J/\psi\phi$ decays is discussed. The performance of this selection is crucial for an experiment in a hadron collider, since there are many sources of background. First, the data are divided into several categories. Then, the specific problem is addressed of how to distinguish the decay type of interest from all sorts of other B-decays. Various parameters are presented, which allow for the separation between signal and background events. The method to optimize these selection criteria is discussed in detail and the results of this optimization are presented.

In the last section, the performance of the experiment is evaluated with simulated data. These data are produced and acquired according to the specifications presented in the previous chapters. The impact of the $B_s \to J/\psi\phi$ measurements of *LHC*b on the physics parameters $\Delta\Gamma/\Gamma$ and $\delta\gamma$ are extracted from the predicted number of signal and background events per year via a toy Monte Carlo study.

6.1 Software

Over the last years, the *LHC*b software has evolved and various changes were made between the large data production run in 2003 (Data Challenge or DC'03) and DC'04. Although the $B_s \to J/\psi\phi$ analysis presented here, is based on data of DC'03, the software discussed is from DC'04. First of all, the *LHC*b concepts for data handling are better implemented in the software of DC'04 and secondly, this avoids discussing obsolete software. In case significant differences between the two versions exist, a remark is made explaining the DC'03 situation.

The software consists of a chain of applications with the following functionality:

- Gauss - Generation of simulated events (in DC'03 this was done by SICB, which also included detector response simulation).

- Boole – Detector response simulation, including the trigger functionality (not available as a separate component in DC'03).

- Brunel – Particle identification and track reconstruction, which allows to determine impact parameters and momenta.

- DaVinci/LoKi – Selection and analysis, which provides invariant mass reconstruction, vertex information, *etc*.

Simulated events are treated in the same way as real data to ensure the applications can handle real data as well. This puts design constraints on the software, to ensure that simulation and reconstruction are completely independent. Various groups in *LHC*b participate in the coding of these applications, making modular development an essential requirement. The organizational aspects and coding conventions are accommodated by a framework named Gaudi. This software framework is presented in section 6.1.1, followed by a discussion of Gauss, Boole, Brunel and DaVinci/LoKi.

6.1.1 Gaudi

Gaudi is an object-oriented framework, written in C++, developed for event data processing applications[1]. It provides a software skeleton with the most common functionalities and allows developers to plug in their task specific code. This framework provides the *LHC*b software developers with standards, *e.g.* a common vocabulary, which facilitates integration of components and thereby enables concurrent development. Besides organizational benefits, a common framework adds robustness and change-tolerance to a program by enabling the re-use of code.

Gaudi is centered around the concept of algorithms. These algorithms perform data manipulations on event by event basis, to complete a (required) task. A number of those manipulations are common and therefore provided by standard services. Examples of such services are: event data service, message service, histogram service, *etc*.

The various components of the *LHC*b software are integrated in Gaudi with abstract interfaces. These interfaces are independent of the concrete implementations (much like *e.g.* the buttons on a CD player, which are more or less the same for all brands and models). One of the most interesting interfaces is the one providing interaction between the algorithms and the so-called transient data store. This transient data store acts like a blackboard, minimizing the coupling between algorithms. An algorithm does not need to know how the data it uses have been produced, it only needs to know what data it requires as input and what data it will be producing as output. This mechanism accommodates the creation of 'split' programs, providing the required modularity. As is shown in Fig. 6.1, data can be processed by scheduling the algorithms in the adequate order, as long as these algorithms know their own inputs and outputs. The machinery behind the abstract interfaces arranges the redirection from apparent to actual data flow.

[1] http://proj-gaudi.web.cern.ch/proj-gaudi

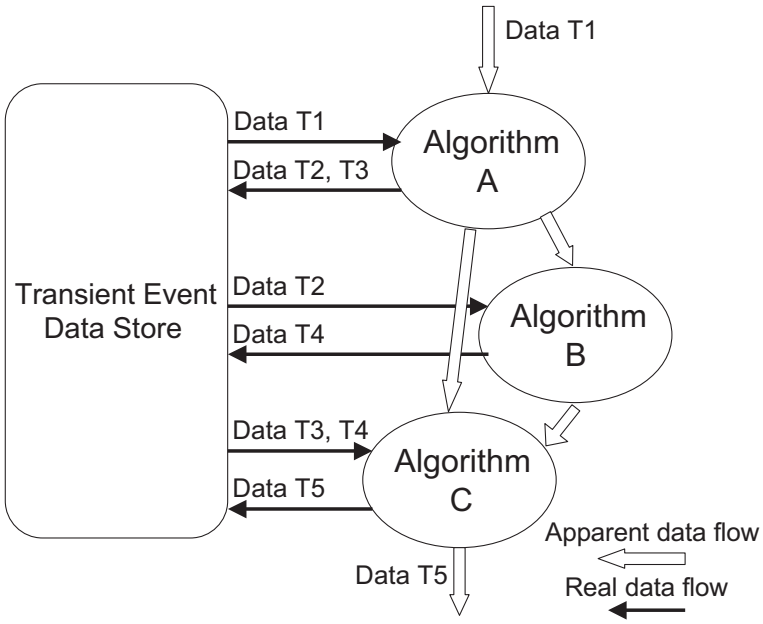


Figure 6.1: Data processing in Gaudi. The open arrows indicate the apparent flow of data, where the closed arrows demonstrate the actual data flow based on the 'blackboard approach'.

6.1.2 Gauss and Boole

The simulation package contains three main components: event generation, detector response simulation and digitization, as shown in Fig. 6.2 [77]. These three components are implemented as algorithms and interact with the transient event data store as discussed in the previous section. The event generation and detector simulation algorithms are grouped together in the Gauss application, where the digitization is done in a separate program called Boole.

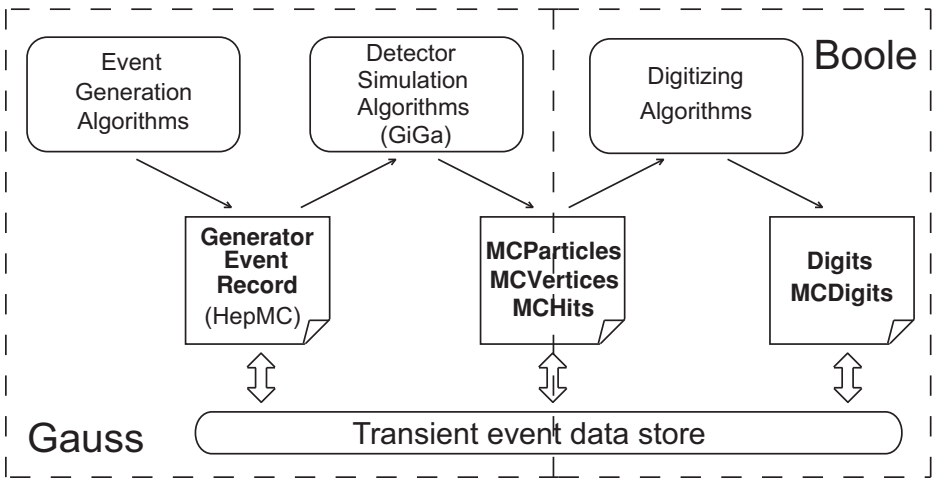


Figure 6.2: Components of Gauss and Boole. The dashed lines separate the two applications.

Simulated events are generated by the PYTHIA event generator, as described in sec-

tion 3.1.1. The decay of B hadrons is calculated by a dedicated package called EvtGen (in DC'03 the package QQ was used), as mentioned in section 3.1.2. These packages are wrapped inside algorithms, to integrate them with the Gaudi framework. The final step of the event generation process consists of an algorithm that takes smearing of the position of the primary vertex into account, which is caused by the particle distributions in the colliding beams.

The next component contains the detector simulation algorithms, which are based on a flexible integration of Geant4[2] into Gaudi (Geant3 in DC'03). The coupling between these two modules is realized via a set of abstract interfaces that form a facade, called GiGa (Gaudi interface to Geant application) [78], around the simulation engine (in DC'03 Geant3 was integrated in SICB). One of the functionalities offered by these interfaces is the access of Geant4 to the transient event data store. Geant4 loads the events created by the event generation algorithms, as well as the *LHC*b geometry model and the magnetic field map. The B decays as simulated by EvtGen are reevaluated by Geant4, since EvtGen does not take into account that particles are affected by the materials they traverse. If necessary, these decays are modified to incorporate the interactions with the detector. Note that the GiGa run manager, instead of implementing the standard black-box approach via the Geant4 run manager, provides internal access to the methods in the Geant4 event loop. As a result, all Geant4 actions are instantiated by GiGa and can be configured when the application is started, just by changing the job options[3]. This flexibility and the resulting unification with the configuration of the rest of the *LHC*b algorithms is probably one of the most interesting features offered by GiGa.

The last component of the simulation is the digitization, simulating the responses to hits in the detector as generated by Geant4 in the previous step. The digitization step includes simulation of the detector response and of the read-out electronics, as well as of the level-0 and level-1 trigger. Boole includes features like the simulation of the effects of spill-over and detector noise. The simulated detector responses are tuned by the sub-detector groups with information obtained from *e.g.* beam tests.

6.1.3 Brunel

Brunel performs the offline reconstruction of *LHC*b events. It contains an implementation of the track reconstruction and particle identification as discussed in section 3.2.2 and 3.2.3, respectively.

Simulated and real data are treated identically in the reconstruction. The *LHC*b event data model is specifically designed to avoid any coupling between the reconstruction and the simulation, as shown in Fig. 6.3. Relationships in the event data model follow the arrows, *i.e.* they only exist between objects from adjacent steps in the data processing. This implies that the only direct relation between objects produced by the simulation and objects processed by the reconstruction exists between Digits and MCDigits. The association between these objects is done implicitly via the electronic channel ID, hence providing a clear separation between the reconstruction and the simulation.

[2] http://cern.ch/geant4

[3] The *LHC*b algorithms are configured with job options. Each time the application starts these options are read from a file.

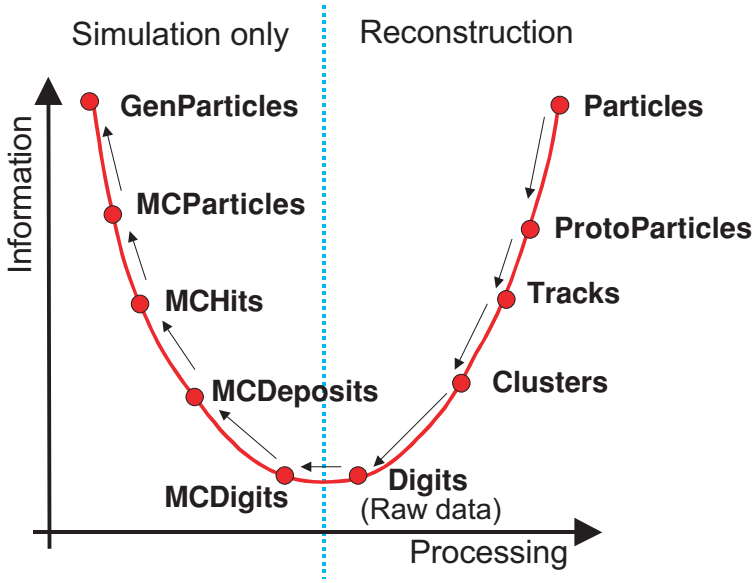


Figure 6.3: Relationships in the event data model. Only backward relations between adjacent objects are directly present. More advanced relations between objects are implemented by means of associator algorithms.

The event data model has far reaching implications for the data analysis routines in *LHC*b. An explicit example is the case where an algorithm needs a relation between distant objects in the data processing chain (*e.g.* reconstructed Tracks to MCParticles). Since there is no direct way to implement this relationship, associator algorithms are required to provide this functionality. These associators can be trivial (*e.g.* the relation between a Digit and a MCHit), or complex, *e.g.* the relation between Tracks and MCParticles, where it is not at all obvious that the reconstructed track indeed contains only hits caused by a single MCParticle. Internally, the associator bases its decision for an association on criteria which can be deduced from explicit links in the data model, but can also be anything else. An associator can provide different kinds of relations, depending on the implementation (*e.g.* one to one, one to many, many to one, many to many or weighted relations). However, the users are shielded from the complexity of this implementation by standard interfaces that provide uniform access to these associators.

6.1.4 DaVinci and LoKi

The end point of the reconstruction and the starting point of the physics analysis is marked by the formation of ProtoParticles. These objects contain all the links to the reconstruction information used to create them, including a list of particle identification hypothesis with their probability. However, they do not have a particular particle identification assigned yet.

The tools offered by DaVinci allow to identify specific decays in the pool of ProtoParticles as generated by the reconstruction. This functionality is, in the spirit of Gaudi, incorporated in a set of basic analysis algorithms. These basic algorithms include tools for

- generation of Particles from ProtoParticles,
- (mass constrained) vertexing and geometrical fits,

- Monte Carlo matching and debugging

These tools can be configured and combined to construct a selection for a particular decay channel. DaVinci in principle offers all functionality needed for such selections. On top of DaVinci, a higher level tool called LoKi is developed. This tool shields the user from a number of technical difficulties and allows them to write code which does not obscure the actual physics content with C++ semantics. The package provides the following functionality:

- Smart loops over particle combinations, preventing double counting.
- Selection and filtering with criteria based on kinematic and topological properties, as well as on particle identification information.
- Access to Monte Carlo truth information.
- Creating and filling histograms and ntuples. In principle this functionality is provided by Gaudi. Remember however, that Gaudi runs the algorithms on an event by event basis. This means that histograms and ntuples need to be created during initialization, filled by the algorithms and saved afterwards. As a result, these objects lose their locality (*i.e.* they are created in one method, filled in another and saved in yet another method), which complicates their handling. LoKi provides a different approach: the algorithm contains a method that instantiates histograms and ntuples the first time it is called. The next time it just returns the existing histogram or ntuple. As a result, handling is contained within the method that runs in the event loop, which allows treatment of these objects as if they are local.

A typical code fragment of LoKi, illustrating the self-explanatory style, is shown in Fig. 6.4.

```
select ( ”muPlus” , ID == ”mu+” && PT > 500/MeV ) ;
select ( ”muMin” , ID == ”mu-” && PT > 500/MeV ) ;
for ( Loop JPsi = loop( ”muPlus muMin” , ”JPsi” ) ; JPsi ; ++JPsi )
{
   plot ( ”Mass of JPsi” , M( JPsi ) / GeV ) ;
}
```

Figure 6.4: LoKi code fragment.

6.2 Event selection

First of all, the number of signal events produced within the *LHC*b detector is calculated for one nominal year of data taking. Next, the possible observables are discussed to distinguish signal from background. The distribution of these observables can be linked to the design specifications on one hand and to the physics performance of the experiment on the other. Finally, the observables are combined with an optimization strategy, which allows to

calculate the best combination of selection criteria based on an optimal ratio of the number of signal events over the number of background events. The section is concluded with the performance numbers of the $B_s \to J/\psi\phi$ selection.

6.2.1 Event production

LHC will produce about 10^{12} $b\bar{b}$ pairs per year, as stated in section 3.1.1. Roughly 10% of the produced b-quarks will form a B_s-meson in the subsequent hadronization process. The branching fraction of $B_s \to J/\psi\phi$ decays is $(9.3 \pm 3.3) \cdot 10^{-4}$ and for the subsequent decays $J/\psi \to \mu^+\mu^-$ and $\phi \to K^+K^-$ it is $(5.9 \pm 0.1) \cdot 10^{-2}$ and $(49.2 \pm 0.7) \cdot 10^{-2}$ respectively[4]. Hence, if both the b and the $\bar{b}$ of the pair are taken into account, 1 in $1.9 \cdot 10^5$ $b\bar{b}$ pairs contains a $B_s \to (J/\psi \to \mu^+\mu^-)\,(\phi \to K^+K^-)$ decay, which results in $10^{12}/1.9 \cdot 10^5 = 5.4 \cdot 10^6$ of these decays per year. Note that about a third of the B-decays has the B-meson of interest contained in the acceptance of *LHC*b, which means about $1.8 \cdot 10^6$ signal events can in principle be reconstructed per year.

The distributed production of Monte Carlo data, as discussed in section 5.1, gave the opportunity to generate 55 million events during the data challenge in 2003. This amount of events still represents only about a second of *LHC*b operation (*i.e.* $55 \cdot 10^6/40\,\text{MHz}$)[5] and would thus, most likely, not contain any signal events. This problem is circumvented by generating specific samples:

- The minimum bias background sample contains about 30 million events. This type of background is not considered here[6]. Reduction of minimum bias background is a general concern (handled by the trigger), not specific to this decay. Instead, we concentrate on reduction of the B-inclusive background, *i.e.* all sorts of events with B-mesons that decay to something else than $J/\psi\phi$.

- The B-inclusive sample contains about 10^7 events. Most of these events are obviously background and will certainly fail to pass the $J/\psi\phi$ selection criteria. It is a clear waste of resources to reconsider these events over and over again during the optimization of our selection criteria. Hence, by stripping the B-inclusive sample once, a subset is created with only the most dangerous background events. The tuning of the selection criteria is based on this subset, which considerably simplifies the data handling.

- The sample with $B_s \to (J/\psi \to \mu^+\mu^-)\,(\phi \to K^+K^-)$ decays contains about 750,000 events, of which 30,000 are used to optimize the selection criteria.

These samples do not represent equal periods of *LHC*b operation. The number of B-inclusive background events, available for the optimization, is only $10^7/30{,}000 \approx 333$ times larger than the number of signal events. This means that, compared to the required $1.9 \cdot 10^5$, the amount of background events needs to be scaled by a factor 570 with respect to the signal

[4] http://pdg.lbl.gov

[5] Note that not all possible bunch crossings of LHC will be filled and some bunches will cross without interactions, so this number provides an indication only. Furthermore, the simulated events did not contain spill over (in DC'03 data the spill over effects are artificially added at a later stage, where as in DC'04 these effects are included).

[6] Note that spill-over for the b-inclusive and signal events is generated from this event sample.

events in order to match the conditions of *LHC*b operation. The samples can be normalized to the amount of data collected in 1 year, by (an additional) scaling to the expected number of signal events (*i.e.* a factor $1.8 \cdot 10^6/30,000 = 60$). Note that this would not affect the results of the performance studies.

6.2.2 Selection criteria

In this section the various parameters are discussed that allow to distinguish $B_s \rightarrow J/\psi\phi$ decays from background events. The most dangerous background consists of other B-decays and combinatorial background (*i.e.* random combinations of tracks that lead to a B-meson candidate).

The excellent track reconstruction and particle identification features of the *LHC*b detector offer a variety of selection criteria to distinguish signal from background. Here, reconstructed mass, (transverse) momentum of the decay products, normalized χ^2 of the vertex fit, particle identification and proper lifetime of the B-meson are discussed. The samples of signal and background Monte Carlo data are preprocessed with loose cuts (see Table 6.1) on these observables, accepting all signal and rejecting only the most obvious background. Later in this section the exact definition of these observables is discussed.

Table 6.1: Pre-selection values for $B_s \rightarrow (J/\psi \rightarrow \mu^+\mu^-)\,(\phi \rightarrow K^+K^-)$. The likelihood of the particle hypothesis i is indicated with $\mathcal{L}_i$. The difference between the reconstructed mass and the PDG mass of particle type i is indicated with ΔM_i. In this analysis mass windows are optimized symmetrically around the PDG mass. Hence, only the criteria for the absolute values of these observables are listed. The symbol for the normalized χ^2 of the vertex fit for particle type i is $\overline{\chi}^2_i$.

$ln\left(\frac{\mathcal{L}_\mu}{\mathcal{L}_\pi}\right)$	$ln\left(\frac{\mathcal{L}_K}{\mathcal{L}_\pi}\right)$	$\lvert\Delta M_{J/\psi}\rvert$	$\lvert\Delta M_\phi\rvert$	$\lvert\Delta M_{B_s}\rvert$	$\overline{\chi}^2_{J/\psi}$	$\overline{\chi}^2_\phi$	$\overline{\chi}^2_{B_s}$
> -20	> -20	< 50 MeV/c^2	< 50 MeV/c^2	< 70 MeV/c^2	< 300	< 300	< 300
				< 700 MeV/c^2 (background)			

Next, the distributions of these observables are determined for reconstructed events in the background sample and in the signal sample. An observable that allows for efficient selection of signal and at the same time rejects background is considered a good discriminator and will be used in the next section to optimize the selection. Note that the criteria available in the off-line analysis are not all available for the trigger (due to the CPU time consumption of the algorithms). In section 6.2.4 the trigger efficiency is calculated from the results obtained here.

Mass windows

One of the most common selection criteria to distinguish between signal and background is the reconstructed mass. In this decay, selection based on the reconstructed mass is especially powerful, since the J/ψ and the ϕ are narrow resonances. Here, mass windows are optimized independently from the other selections for a number of reasons:

- Preferably, all selections should be optimized together. However, this is not possible due to CPU time consumption (see section 6.2.3).

- As discussed in the previous section, there are not enough simulated background events available in comparison to simulated signal events. Furthermore, the effects of the mass windows on the number of accepted signal and background events are quite predictable. Hence, applying these selections would result in the reduction of background events, but would not add new insights. Instead of rejecting these background events, they can be used to optimize the other selection criteria.
  As a follow up on this argument, the mass window on background events should be opened even further. Accepting as much simulated background events as possible allows for better optimization of the rest of the observables. Therefore, a 10 times larger B_s-mass window is applied for background events than for signal events (which reduces the scale factor between background and signal events to 50).

Note that the proposed approach is only valid when the optimal values for the other selections are not affected by these choices (*i.e.* they should not be correlated to selections based on reconstructed mass). This can become a problem if the B_s mass window is extended beyond resonances formed by similar decays. These decays can form so-called reflections (they can distort the spectrum of the reconstructed mass) due to *e.g.* misidentification of one of the particles. In section 6.2.3 the consequences of opening the mass windows and optimizing them separately are discussed in more detail.

Figure 6.5 shows the mass distributions of signal and background events after applying the preselection, with exception of the selection on the plotted variable. A couple of features are apparent. The $\phi \rightarrow K^+K^-$ decay channels opens at 988 MeV, two times the invariant mass of the kaon. The cut-off in the J/ψ mass spectrum, on the other hand, is due to a preselection cut on the dimuon invariant mass. The decay channel $B_s \rightarrow J/\psi\phi$ opens at 4.1 GeV, generating a peak in the (combinatorial) background spectrum around 4.2 GeV. The B_s mass is reconstructed directly from the invariant masses and momenta of 2 kaons and 2 muons without constraining them to the J/ψ or ϕ mass, which is clearly demonstrated by the continuing spectrum well below the minimal possible reconstructed invariant mass of a $J/\psi\phi$ combination. In the background sample, the fraction of J/ψ and ϕ events is 0.4% and 0.1% respectively, which does not result in visible J/ψ and ϕ signals in these plots.

Transverse momentum of final state particles

The average transverse momenta of the B_s decay products are larger than the transverse momenta of background particles, due to the relatively large B_s mass. This discriminator is implemented as a selection based on the transverse momentum of the final state particles. The decay will only be marked as signal when the transverse momenta of at least one of the kaons and at least one of the muons exceed the values determined in the optimization (section 6.2.3).

Momentum of ϕ-mesons

In Fig. 6.6 the distribution of the momenta of the ϕ-mesons is shown for signal and background. The average momentum in the background events is lower than that of the sig-

nal events, which indicates that heavy quarks are produced with higher momenta than light quarks (*i.e.* the momentum of a heavy quark does not change much when it recombines with a light quark). Qualitatively, one could think of several arguments for the differences in the momentum distribution:

- In a hadron collider, much more interactions take place with low than with high interaction energy (see Fig. 6.7 [79]). In these low energy interactions the production of light flavors is favored.

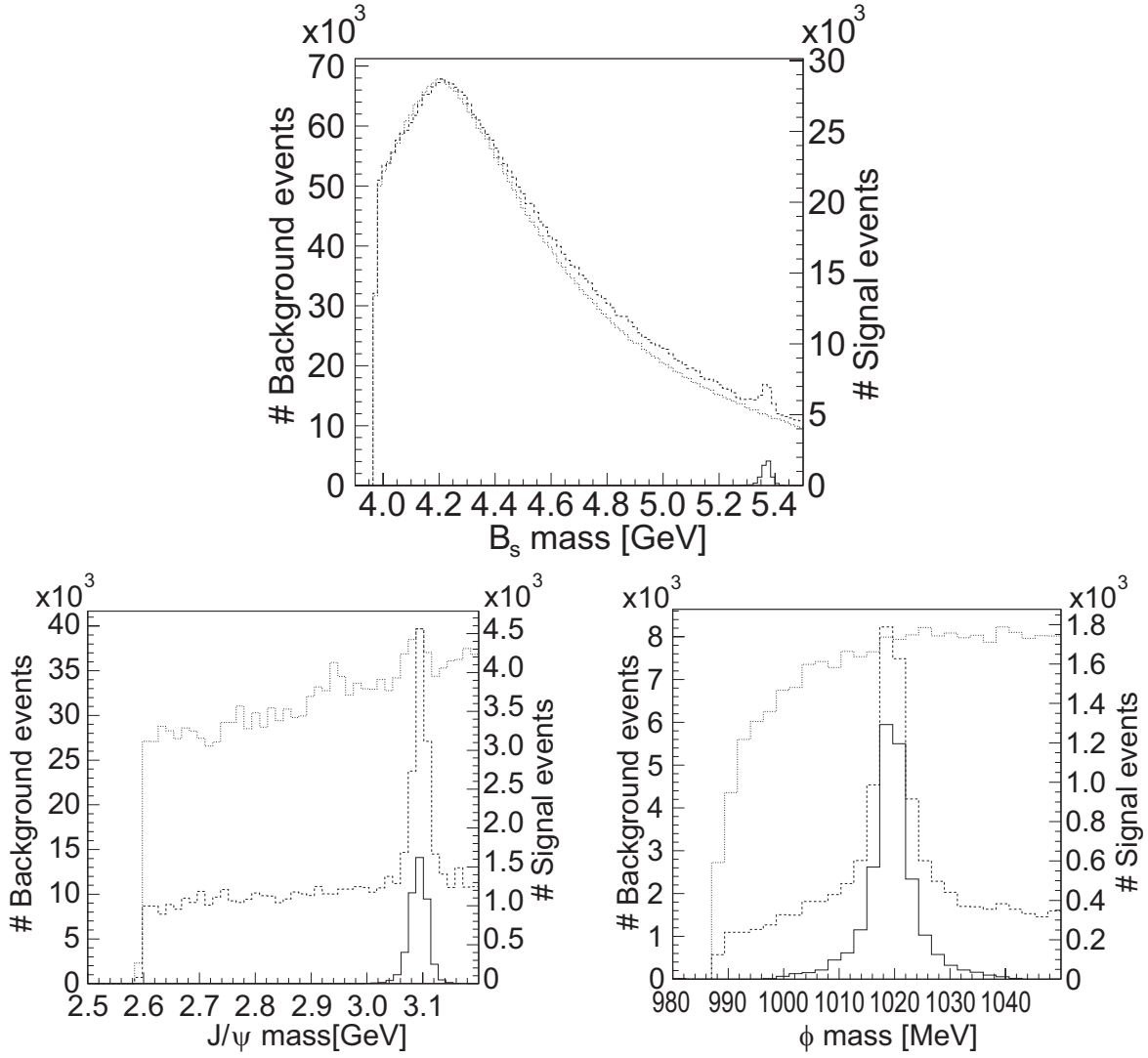


Figure 6.5: Mass distributions of signal and background events. The solid lines represent the Monte Carlo truth information of particles involved in a $B_s \rightarrow J/\psi\phi$ decay, the dashed lines are reconstructed signal (including combinatorial background from this signal), the dotted lines show the number of background events.

- More light than heavy flavors are produced in cascaded production processes. This can be understood by considering high energy gluons losing part of their energy to *e.g.*

quarks produced via flavor excitation. Hence, even if sufficient energy is available to neglect the quark masses at first (*i.e.* heavy and light flavours will be produced in equal amounts), further down the cascade only light flavours will be produced.

- In the Lund model, used by Pythia, the dominant production process for light flavours is based on string fragmentation. In string fragmentation models, interaction energies are distributed over regions in space time enclosed by the interacting particles. The decay products of these strings each carry part of the original interaction energy. In [80] it is concluded that the energy densities generated by these processes are insufficient for the production of heavy flavours.

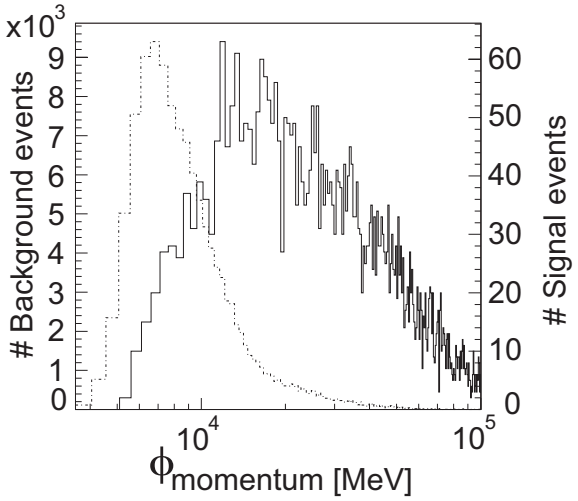


Figure 6.6: Distribution of ϕ momenta for signal (solid line) and background events (dashed line). The bin width is 0.5 GeV/c. Note the logarithmic x-scale.

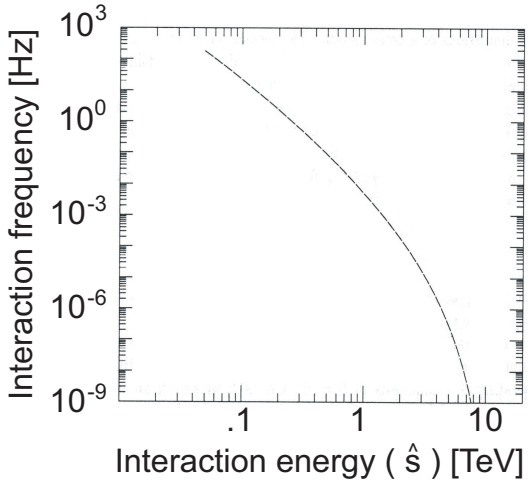


Figure 6.7: Relation between the quark−gluon interaction frequency and the interaction energy ($\hat{s}$) for LHC experiments (at $10^{32} cm^{-2} \cdot s^{-1}$).

These arguments lead to the conclusion that *b*-quarks are predominantly produced in interactions with a relatively high energy. As a result, the gluons involved in these interactions cover a broader momentum range and allow on average for a larger momentum asymmetry. This explains the differences in the momentum distribution of promptly produced ϕ-mesons versus those originating from *B*-meson decays, which leads to the differences in the distributions of signal and background samples as shown in Fig. 6.6. Hence, only events with a ϕ which has sufficient momentum will be accepted.

Normalized χ^2 of the decay vertices

Three decay vertices can be identified in this cascaded decay. The excellent resolution of the vertex detector will provide a powerful parameter to distinguish between actual vertices and combinatorial background. If the opening angle of the decay products is sufficient, the normalized χ^2 $\left(\equiv \overline{\chi^2}\right)$ of the fit that tries to reconstruct these vertices can accurately be determined. A large $\overline{\chi^2}$ is a strong indication that the tracks involved in the fit do not originate

from a single vertex. These track combinations can then safely be marked as combinatorial background.

There are two cases where the combination of random tracks still results in an acceptable vertex fit. The first case is when both particles originate from the primary vertex, in which case the proper lifetime significance (see further in this section) can be used to identify them as prompt background. The second case is when particle trajectories coincidentally intersect. In this case other parameters need to provide the distinction between signal and background. Note that the decay of a high momentum $\phi(1020\,\mathrm{MeV/c^2})$ into two kaons $(2\cdot 494\,\mathrm{MeV/c^2})$ results in a vertex with a small opening angle, which is difficult to reconstruct. As a result, the $\overline{\chi^2}$ of the ϕ vertex fit does not provide sufficient separation between signal and background events.

Particle identification

Particle identification is based on likelihood ratios as described in section 3.2.3. The best value for selection based on this criterion is difficult to determine. Part of the background is formed by real kaons and muons, often coming from real J/ψ and ϕ particles. Hence, particle identification is not useful in removing this type of background. This problem demonstrates itself by strong correlations between the optimal log likelihood selection value and *e.g.* selection based on the invariant mass of the J/ψ and ϕ. In order to optimize the selections based on log-likelihood, background with real kaons and muons needs to be removed first. This is demonstrated in Fig. 6.8, where the log-likelihood distributions are shown for background events that passed the preselection with and without kaons and muons. The separation of muon and non-muon background is not as clear as expected. This could be caused by *e.g.* pions which decay in flight to muons. If these particles are identified as non-muon background by the Monte Carlo association, they will pollute the spectra and blur the separation of the two categories.

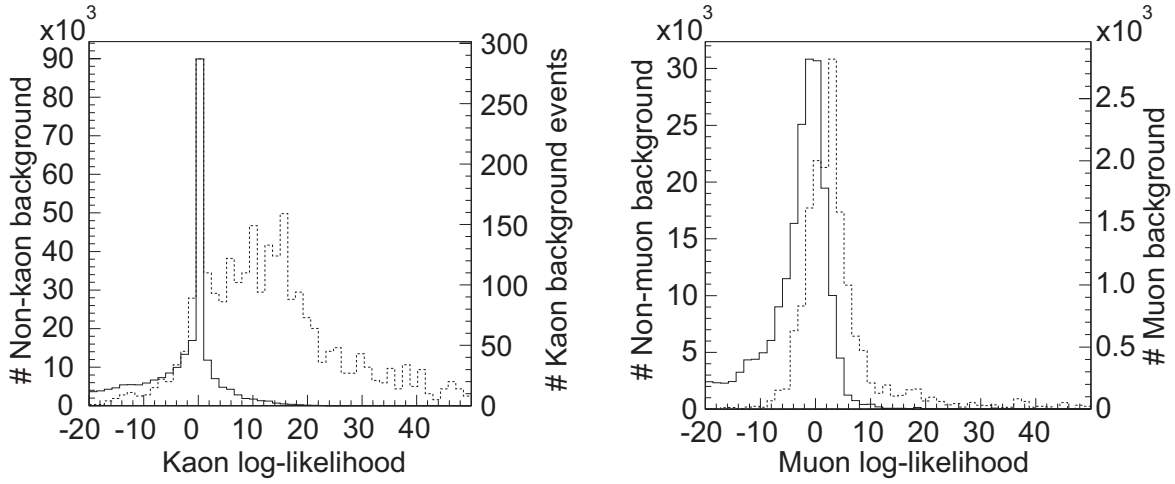


Figure 6.8: Log-likelihood distributions for background events with and without kaons (left) and muons (right).

Note that the opening of mass windows, as discussed earlier, can introduce 'fake' background (*i.e.* non typical background caused by other resonances), affecting the optimization

of selections based on particle identification.

Proper lifetime significance of the B-meson

$B_{(s)}$-mesons have a relatively long proper lifetime. Due to their high boost (see section 3.1.1), this value can accurately be reconstructed from the decay length. This offers the opportunity to distinguish them from other particles and drastically reduces the large combinatorial background of promptly produced J/ψ, ϕ, μ or K. The selection criterion is defined as the proper lifetime divided by the calculated error on this parameter. A description of this calculation is given in Ref. [81]. Whenever this fit is made for a combination of tracks that does not originate from a B decay, this results either in a small proper lifetime (for prompt background) or a large error on the parameter and/or a large χ^2 of the proper lifetime fit. Thus, combined with the χ^2, this selection criterion filters both on the reconstructed proper lifetime and on the quality of the proper lifetime fit.

6.2.3 Optimization.

The observables defined in the previous section can all be used to discriminate between signal and background events. However, none of them is perfect. The optimal performance is the result of a combination of selection criteria[7]. Here, 'optimal' is defined as the best ratio of $Signal/\sqrt{Signal+Background}$, which is consistent with finding the maximum significant observation of the signal events in the Monte Carlo simulation (*i.e.* the interesting quantity is the number of signal events and the statistical error in a bin is calculated as the square root of its entries). Note that any other ranking mechanism can be used as long as it provides a performance number for each combination of selection parameters. Scaling the signal and background simulated event samples and correcting for the mass window method, as discussed in section 6.2.1 and 6.2.2 respectively, results in a redefinition of the ratio to $\left(\sqrt{60}\right) Signal/\sqrt{Signal+57\cdot Background}$ for the samples considered.

Ideally, one would require independent or orthogonal selection criteria, *i.e.* changing the value of one selection criterion does not affect the optimal values of the others. Note that the possibility of measuring observables independently, does not guarantee this requirement is fulfilled. Most of the selection criteria are correlated in one way or another and can therefore only be optimized accurately when they are combined. In order to optimize correlated selection criteria, boundaries are defined for each possible selection parameter based on common sense. This parameter space is scanned to test how many signal and background events pass the selection criteria. This allows us to select the combination of parameters with the best performance.

A practical complication limits the possibility of scanning the whole parameter space. To get a performance number for a set of parameter values, one needs to test all events. If a scan is made of the parameter space, the number of calculations to perform is proportional to the number of parameter combinations multiplied with the number of events used. The CPU time consumption for such an analysis thus roughly scales with the power of the number of correlated observables.

[7] One could even consider to exploit dependencies between selection criteria with other tools, like neural networks.

By applying a counting procedure the number of calculations can drastically be reduced. The data are sorted once for each observable, creating an N-dimensional view of the data (with N the number of observables). Instead of testing all events for every set of parameters, only events that survived the criteria applied in the previous step need to be tested. As soon as an event fails to pass the selection, the counting algorithm moves to the next combination of parameters. A simple example with 5 events and 2 observables is shown in Fig. 6.9. The markers indicate events in the parameter space. The lines represent combinations of selection parameters. The numbers indicate how many events passed the selection. A variable bin size method is applied, where each bin edge matches the coordinates of an event.

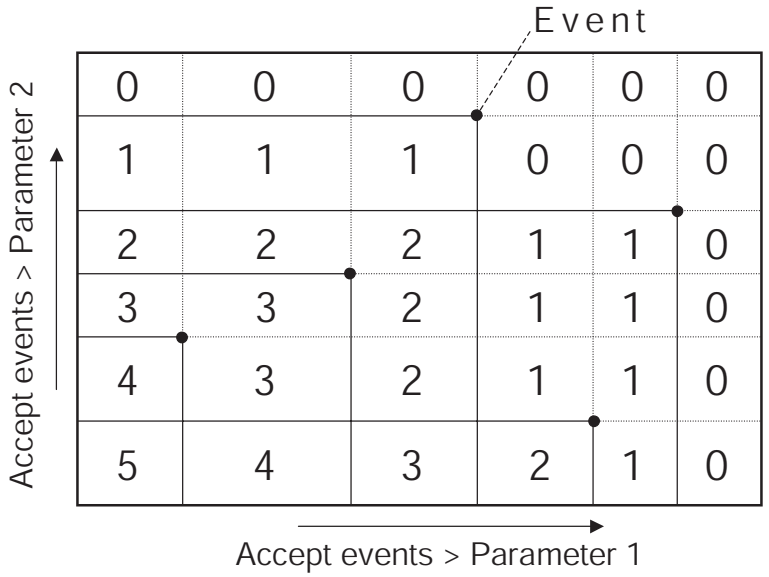


Figure 6.9: Counting procedure, based on an N*-dimensional view on the data. The events are marked as black dots. Whenever one of the solid boundaries is crossed, either in the direction of Parameter 1 or in the direction of Parameter 2, an event is lossed because it no longer passes the criteria. Like this, each event is only tested once, instead of testing all events for all parameter combinations.*

A plot like this can be created for signal and background events, resulting in two N-dimensional arrays like Fig. 6.9 with either fixed or variable bin sizes. The total number of array elements equals the number of tested parameter combinations. The performance numbers can be calculated by combining the information from these two arrays. On a 2 *GHz* machine the analysis of about 100,000 events, with 5 correlated selection criteria (20 bins each) takes roughly 45 minutes and uses a few hundred *Mb* of memory.

The $B_s \rightarrow J/\psi\phi$ selection contains more than 5 criteria and even with the counting algorithm it still poses a problem with respect to CPU time consumption. A possible solution for this problem is a chained analysis:

- The first step is to discriminate between reasonable candidates for this decay and (combinatorial) background, based on the decay topology. This topology can be distinguished by the $\overline{\chi}^2$ of the B_s and J/ψ decay vertices, together with the momentum of the ϕ and the transverse momenta of the kaons and muons. The results are $\overline{\chi}^2_{B_s} < 3.5$,

$\overline{\chi}^2_{J/\psi} < 3$, $p_\phi > 20\,\text{GeV/c}$, $\max\left(p_{T,\mu^+}, p_{T,\mu^-}\right) > 2300\,\text{MeV/c}$, $\max\left(p_{T,K^+}, p_{T,K^-}\right) > 1000\,\text{MeV/c}$.
Note that the relation between the signal significance and $\overline{\chi}^2_{J/\psi}$ is almost flat in the range from 3 to 10 and uncorrelated with the $\overline{\chi}^2_{B_s}$ (see Fig. 6.10) . A conservative approach would be to accept events with larger $\overline{\chi}^2$ at first (*e.g.* up to 10), to make sure (based on real data) that a tighter selection does indeed not affect the significance.

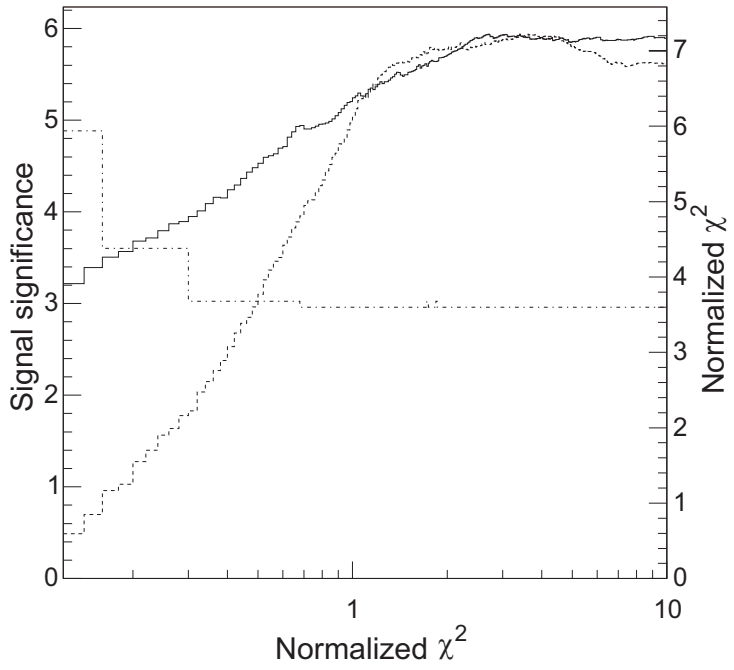


Figure 6.10: The solid (dashed) curve shows the relation between the signal significance and $\overline{\chi}^2_{J/\psi}$ $\left(\overline{\chi}^2_{B_s}\right)$. The dash dotted line shows the correlation between $\overline{\chi}^2_{J/\psi}$ (x-axis) and $\overline{\chi}^2_{B_s}$ (y-axis). Note the logarithmic x-scale.

- The optimal values determined in the first step are applied as pre-selection criteria on the Monte Carlo events. The remaining events provide J/ψ and ϕ candidates which could come from B_s decays. Still, two background categories are not addressed properly:

  - Promptly produced J/ψ and ϕ particles, which appear to form a B_s. This category can be identified by the proper lifetime significance. Selecting only events with $\tau_{B_s}/\sigma_\tau > 6.5$, would result in the best signal significance. However, in the analysis of section 6.3 this proper lifetime significance is not implemented as a selection criterion. Instead, the shapes of signal and background as a function of proper lifetime act as input for the likelihood fit.
  - Mis-identified $B_s \to J/\psi\phi$ decays and/or decays of B mesons to something else then $J/\psi\phi$, but with a similar topology (*e.g.* $B_d \to (J/\psi \to \mu^+\mu^-)\,(K^* \to K^+\pi^-)$).

These decays can be rejected based on particle identification. However, the selection criteria for particle identification are strongly correlated with the mass windows, which are not yet determined. Moreover, part of the background still contains real muons and kaons. First, the remaining background is filtered to reject these events and to avoid pollution of the optimization results. Next, the optimal selection for the log likelihood is expressed as a function of the corresponding mass windows. The relation between mass window and log-likelihood is approximated with a linear equation in the region of interest. The resolution is insufficient for a higher order approximation due to the relatively small number of available background events. The resulting requirements are $\Delta ln\left(\mathcal{L}_{K-\pi}\right) > -4.5 + 0.3 \cdot \left|\Delta M_{\phi}\right|$ for the separation between kaons and pions and $\Delta ln\left(\mathcal{L}_{\mu-\pi}\right) > -4.5 + 0.1 \cdot \left|\Delta M_{J/\psi}\right|$ for the separation between muons and pions.

- As mentioned in the previous section, mass windows are optimized independently from the rest of the selections. The main objective of the B_s mass window is elimination of random combinations of J/ψ and ϕ particles that appear to come from a common vertex. The reconstructed mass of these combinations has no bias toward the B_s mass. Hence, the other selections are expected to be uncorrelated with a mass window on the B_s. The mass windows on the J/ψ and ϕ are expected to have correlations with the other selection criteria.
  The optimal values, which are found for these criteria are $|\Delta M_{B_s}| < 17.5\,\mathrm{MeV/c^2}$, $\left|\Delta M_{J/\psi}\right| < 13.5\,\mathrm{MeV/c^2}$, $\left|\Delta M_{\phi}\right| < 3.5\,\mathrm{MeV/c^2}$. They can be compared with the estimate of 2 times the width of the distributions found in the Monte Carlo signal data (*i.e.* about 95% of the signal remains), which gives $|\Delta M_{B_s}| < 33\,\mathrm{MeV/c^2}$, $\left|\Delta M_{J/\psi}\right| < 25\,\mathrm{MeV/c^2}$, $\left|\Delta M_{\phi}\right| < 5.5\,\mathrm{MeV/c^2}$. These conditions are more relaxed than the ones found with the optimization, which is an indication that correlations indeed exists.
  As an illustration of the encountered problems with the optimization, the ϕ mass window is discussed. The ϕ resonance has a width of about 4.3 MeV and apparently this narrow width offers an efficient way to reduce the enormous background with an extremely tight mass window. However, as shown on the left hand side in Fig. 6.11, combination with other selection criteria (in this case the transverse momentum of the kaons), allows to relax the mass window again. Hence, it should not be optimized independent from the other criteria. The same statement can be made for the mass window of the J/ψ, of which the most pronounced correlations are shown on the right hand side of Fig. 6.11. Note that the values found for the significance in these correlation plots are meaningless, since they are calculated only in combination with the mass windows[8].

- There are a number of possible solutions for these problems. Of course, the availability of more background data is the first requirement. In a running experiment, these background events can be obtained from the side bands of the B_s mass spectrum. However, the optimization would require a significant amount of CPU time if the algorithm needs to process a lot more background data than has been done for this analysis.

[8] This could be solved by applying all other selection criteria before the correlations are calculated. Unfortunately, the number of remaining background events after such a pre-selection is insufficient to allow making these plots.

The processing time can be reduced by combining some of the selection criteria, *e.g.* by performing mass constrained vertex fits. Furthermore, it is straightforward to parallelize the algorithm and distribute it over hundreds or thousands of machines (*e.g.* by using grid computing). A third option is to split the optimization in smaller pieces and iteratively determine the optimum. This approach is more sensitive for local optima, but much faster since only few parameter combinations are calculated.

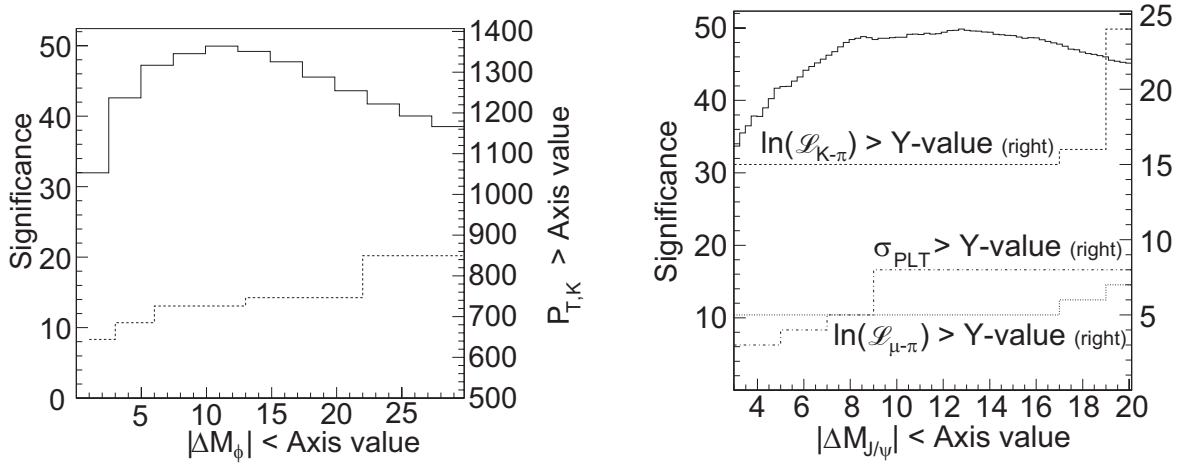


Figure 6.11: Correlation plots of J/ψ and ϕ mass windows. The solid curves are the relations between mass windows and signal significance. These curves use the Y-axes on the left hand side of the figures. The other curves show correlations with other selection criteria. These curves use the Y-axes on the right hand side.

6.2.4 Efficiency and purity

The results of the optimization are summarized in Table 6.2. With these settings, no background events are selected. The sample is estimated to be better than 89% pure with 90% confidence in the estimate. The problem, as expected, is the efficiency of 3%.

In order to improve on these results, correlations with mass windows should be taken into account, which would relax the selection criteria. As a motivating example, the results are compared to those of Ref. [81], as shown in the second column of Table 6.2. In this scenario the efficiency is $\sim 10\%$ and 10 background events are accepted, resulting in a purity of about $71-88\%$ with 90% confidence in the estimate. The same efficiency can be reached with only 4 background events (*i.e.* $85-97\%$ pure with 90% confidence in the estimate), by relaxing some of the selection criteria presented here. These relaxed criteria are shown in the third column of Table 6.2. In Fig. 6.12 the mass distributions for signal and background are shown after applying these cuts.

The events that passed the relaxed criteria served as input for the calculation of the trigger efficiency. The level-0 trigger accepted 87% of these events and subsequently the level-1 trigger accepted 69% from the events that passed the level-0 trigger.

6.3 Sensitivity studies

The performance of an experiment can be estimated with so-called Toy Monte Carlo studies. In section 6.3.1 the principles of these kind of studies are briefly discussed, together with some of the specific implementations required for the $B_s \rightarrow J/\psi\phi$ analysis. The technical details of the implementation of the model are described in Ref. [82]. The code to generate and fit the events is based on the RooFit toolkit for data modeling [83][9] and was also used for the optimization studies of the $B_s \rightarrow J/\psi\phi$ performance as presented in Ref. [2]. This chapter ends with an update of the *LHC*b performance numbers, based on the new results for selection efficiency and purity as found in the previous section.

Table 6.2: Results of the optimization of the signal significance.

	Optimized	From ref. [81]	Retuned
$\overline{\chi}^2_{J/\psi}$	< 10	< 9	< 10
$\overline{\chi}^2_{B_s}$	< 3.5	< 4	< 5
$\max\left(p_{T,\mu^+}, p_{T,\mu^-}\right)$	$> 2300\,\mathrm{MeV/c}$	–	$> 1000\,\mathrm{MeV/c}$
$\min\left(p_{T,\mu^+}, p_{T,\mu^-}\right)$	–	$> 500\,\mathrm{MeV/c}$	–
$\max\left(p_{T,K^+}, p_{T,K^-}\right)$	$> 1000\,\mathrm{MeV/c}$	–	$> 500\,\mathrm{MeV/c}$
$\min\left(p_{T,K^+}, p_{T,K^-}\right)$	–	$> 500\,\mathrm{MeV/c}$	–
p_ϕ	$> 20\,\mathrm{GeV/c}$	$> 12\,\mathrm{GeV/c}$	$> 10\,\mathrm{GeV/c}$
$\max\left(ln\left(\frac{\mathcal{L}_{\mu^+}}{\mathcal{L}_\pi}\right), ln\left(\frac{\mathcal{L}_{\mu^-}}{\mathcal{L}_\pi}\right)\right)$	> -2	–	> -2
$\min\left(ln\left(\frac{\mathcal{L}_{\mu^+}}{\mathcal{L}_\pi}\right), ln\left(\frac{\mathcal{L}_{\mu^-}}{\mathcal{L}_\pi}\right)\right)$	–	> -20	–
$\max\left(ln\left(\frac{\mathcal{L}_{K^+}}{\mathcal{L}_\pi}\right), ln\left(\frac{\mathcal{L}_{K^-}}{\mathcal{L}_\pi}\right)\right)$	> -3	–	> -3
$\min\left(ln\left(\frac{\mathcal{L}_{K^+}}{\mathcal{L}_\pi}\right), ln\left(\frac{\mathcal{L}_{K^-}}{\mathcal{L}_\pi}\right)\right)$	–	> 0	–
$\lvert\Delta M_\phi\rvert$	$> 5.5\,\mathrm{MeV/c^2}$	$> 20\,\mathrm{MeV/c^2}$	$> 7\,\mathrm{MeV/c^2}$
$\lvert\Delta M_{J/\psi}\rvert$	$> 25\,\mathrm{MeV/c^2}$	$> 50\,\mathrm{MeV/c^2}$	$> 30\,\mathrm{MeV/c^2}$
$\lvert\Delta M_{B_s}\rvert$	$> 33\,\mathrm{MeV/c^2}$	$> 50\,\mathrm{MeV/c^2}$	$> 40\,\mathrm{MeV/c^2}$
$\tau_{B_s}/\sigma_\tau > 5$	> 6.5	> 5	> 6.5

Signal events		Background events		Performance	
Accepted	Rejected	Accepted	Rejected	Efficiency	Purity
1141	28.859	0	10.000.000	$\sim 3\%$	$> 89\%$ @ 90% CL

6.3.1 Toy Monte Carlo simulations

In the toy Monte Carlo simulation, multidimensional probability density functions (PDFs) are constructed which are expected to mimic the spectra that will come from an analysis of data acquired by *LHC*b. Hence, first one needs to describe the behavior of the experiment with PDFs. The observables of interest for this analysis are B_s proper life time, angular

[9] http://roofit.sourceforge.net

distribution of the muons and the B_s-mass distribution. Next, events are generated according to these PDFs and finally a likelihood fit to the generated data is performed, to extract the parameters and their covariance matrices.

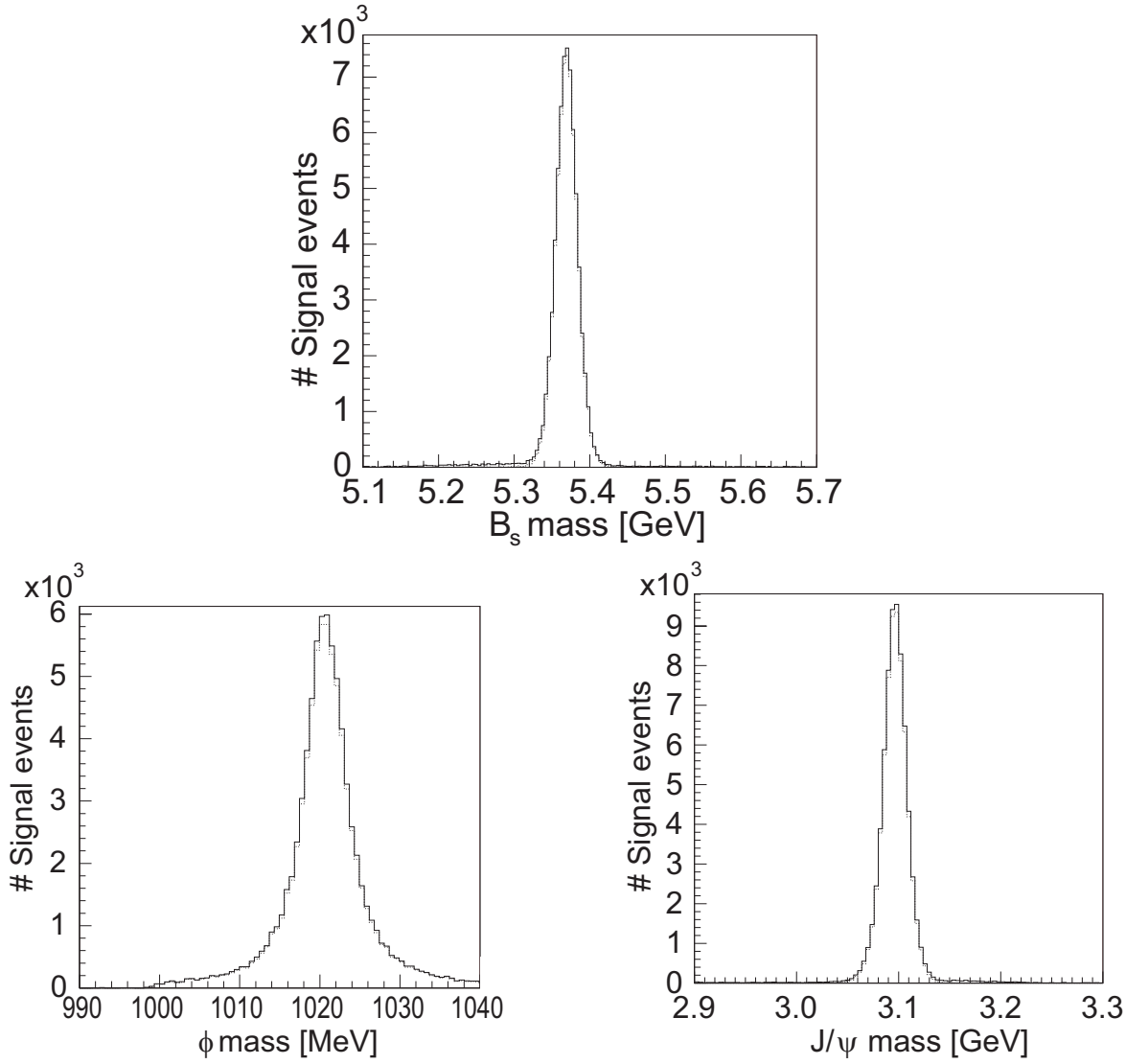


Figure 6.12: Optimized mass distributions. The solid lines are the reconstructed signals and the dotted histograms represent the Monte Carlo truth information.

B_s proper lifetime

The expected proper lifetime distribution of the B_{heavy} and the B_{light} is given by equation (2.13). Experimentally, the oscillations are diluted with the mis-tag rate w, which indicates what fraction of the initial particles is wrongly identified as antiparticle and vice versa.

$$\begin{aligned}
f_{B_s,even}(t) &= (1-R_T)(1-2w)\left(e^{-\Gamma_{light}\cdot t} + e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right)\\
f_{B_s,odd}(t) &= R_T(1-2w)\left(e^{-\Gamma_{heavy}\cdot t} - e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right)\\
f_{\overline{B_s},even}(t) &= (1-R_T)(1-2w)\left(e^{-\Gamma_{light}\cdot t} - e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right)\\
f_{\overline{B_s},odd}(t) &= R_T(1-2w)\left(e^{-\Gamma_{heavy}\cdot t} + e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right).
\end{aligned} \tag{6.1}$$

Besides mis-tagging, the measured proper life time (t_m)contains a double Gaussian resolution function, which needs to be convoluted with the theoretically predicted proper life time distribution, shown in equation (6.2).

$$R_{sig}(t_m - t) = f_1 G(t_m - t; S_1\sigma_t, \mu_1\sigma_t) + (1-f_1)\,G(t_m - t; S_2\sigma_t, \mu_2\sigma_t). \tag{6.2}$$

The core Gaussian describes normal resolution effects, where the second Gaussian provides collects events with distorted results due to *e.g.* a partially mis-reconstructed track, a mis-aligned station, or multiple scattering. The Gaussians of the resolution function scale with the proper lifetime error (σ_t). In other words, the per event error on the proper life time fit provides an indication for the resolution of the measurement. The resulting likelihood distribution becomes

$$\begin{aligned}
R_{sig}(t_m - t) \otimes f_{B_s}(t) = \quad & \left(f_1 G(t_m - t; S_1\sigma_t, \mu_1\sigma_t) + (1-f_1)\,G(t_m - t; S_2\sigma_t, \mu_2\sigma_t)\right) \otimes \\
& \left(p(0)(1-2w)\left(e^{-\Gamma_{light}\cdot t} + e^{-\overline{\Gamma}\cdot t}\sin(2\cdot\delta\gamma)\sin(\Delta m\cdot t)\right)\right).
\end{aligned} \tag{6.3}$$

This parametrization is checked on fully reconstructed Monte Carlo data. The values of the parameters f_1, S_1, μ_1, S_2 and μ_2 can be extracted from a fit to these data (or from a fit to real data). Predictions of the parameters $\delta\gamma$, Δm, Γ and R_T are given by theory. The data are generated with an estimated value of $w = 0.35$. When the data are fitted this number will have to be extracted from control samples such as $D_s^{\mp}\pi^{\pm}$ [2].

The expected proper lifetime distribution of background data is described by a delta function at $t = 0$ for the prompt decays and an exponential for misreconstructed B-decays. This model is also convoluted with a double Gaussian. However, the width of the Gaussian is not correlated with the per event error of the proper life time fit in this case. For background events the proper life time fit assumes either the wrong particle hypothesis or has a wrong notion of the vertex, which means that the corresponding errors are not representative of the RMS spread.

$$\begin{aligned}
R_{bg}(t_m - t) \otimes f_{background}(t) = \quad & \left(f_1 G(t_m - t; \sigma_1, \mu_1) + (1-f_1)\,G(t_m - t; \sigma_2, \mu_2)\right) \otimes \\
& \left(f_{prompt}\,\delta(t) + (1 - f_{prompt})\,e^{-t/\tau}/\tau\right).
\end{aligned} \tag{6.4}$$

The nominal values of the parameters f_1, f_{prompt}, τ, σ_1, μ_1, σ_2 and μ_2 are extracted from reconstructed Monte Carlo data again.

Note that one of the selection criteria in the previous section was the proper lifetime significance. This criterion drastically affects the selection efficiency, leaving it at almost zero

for events with a proper life time below 0.2 ps. Although in principle, one could avoid the introduction in the parametrization by selecting only events with a larger proper life time, this would not solve the problem, since the level-1 trigger contains event selection based on impact parameter significance and this affects the selection efficiency as well (up to about 2 ps). As a result, one should either incorporate the exact shape of the efficiency curve, or partially trigger events with an unbiased trigger with respect to proper time. The second solution is preferred, since it is extremely difficult to determine the exact shape of the efficiency curve. However, this would significantly increase the data flow and the consequences are currently under discussion in *LHC*b. This analysis is still based on the (wrong) assumption that the efficiency is zero before 0.2 ps and constant after.

Angular distribution of the muons

The CP even and CP odd states can be separated with a partial angular analysis. Based on equation (2.12)

$$f_{even} \propto 1 + \cos^2 \theta_{\rm tr},$$

$$f_{odd} \propto \sin^2 \theta_{\rm tr} = 1 - \cos^2 \theta_{\rm tr}. \tag{6.5}$$

Here, $\theta_{\rm tr}$ is the angle between the positively charged muon and the kaon decay plane in the rest frame of the J/ψ. The background sample is generated without angular dependence, although this dependence is implemented as a variable when the data are fitted.

In the Monte Carlo signal sample, the reconstructed $\theta_{\rm tr}$ was compared with the true angle. The RMS of the residuals is found to be about 20 mrad. This resolution is not taken into account in the PDF, but is implemented by smearing the generated angles with a Gaussian of equal width.

B_s mass distribution

The mass distribution of the B_s signal events is described with a Gaussian, since the dominant contributions to its shape is the detector resolution. The background is described by an exponential. All information to describe and generate these events can of course be extracted from the Monte Carlo data samples.

Information about the background events is extracted from the sidebands of the B_s mass spectrum, which provides an important consistency check of the analysis of the CP parameters. Note that there are two effects contributing to the significance of the CP asymmetry. The first effect consists of the uncertainty in the CP asymmetry of the background events. A large uncertainty on this value would reduce the sensitivity for a CP asymmetry in the signal. The second part is expected to scale with the signal significance, *i.e.* $Signal/\sqrt{Signal + Background}$ as was optimized in section 6.2.3. This number will be maximal in the peak of the mass distribution, since the background is almost flat.

6.3.2 Results

The final PDF is the product of the functions described in the previous section and describes the distribution of events measured with a typical experiment that complies to the specifications. One thousand data sets were generated with this PDF. Each of these sets represents the data collected after measuring for one nominal year with a 'toy experiment'. Note that these jobs are well suited for submission to a grid infrastructure, as discussed in Chapter 5. They require practically no input data, can be run in parallel and the output consists of a small number of fitted parameters and their covariance matrices. The turn around time between submission and collection of the results was only 2 days. In this period on average 40 jobs ran concurrently.

The physics parameters for these simulations were set to $\delta\gamma = 0.02$, $R_T = 0.2$, $\Delta\Gamma_s/\overline{\Gamma_s} = 0.1$, $\overline{\Gamma_s} = 1.5\,\text{ps}$ and $\Delta m_s = 20\,\text{ps}^{-1}$. On the left hand side of Fig. 6.13 a projection of one of the likelihood fit results on the $\cos\theta_{\text{tr}}$ distribution is shown, for a random data set with a signal sample of 100,000 events and a purity of 87%. The sum as well as the separate contributions of the background, CP even and odd components are shown. On the right hand side a plot is shown of the oscillations related to Δm_s. With nominal parameters, Δm_s needs to be measured with a different decay and provided as an external parameter, since this decay does not have sufficient sensitivity to extract it. Here, for illustration purposes, $\sin(2\delta\gamma)$ is about 10 times its nominal value of 0.02. The other likelihood fit parameters are R_T, $\delta\gamma$, $\Delta\Gamma_s/\overline{\Gamma_s}$, $\overline{\Gamma_s}$, and w (from a control sample).

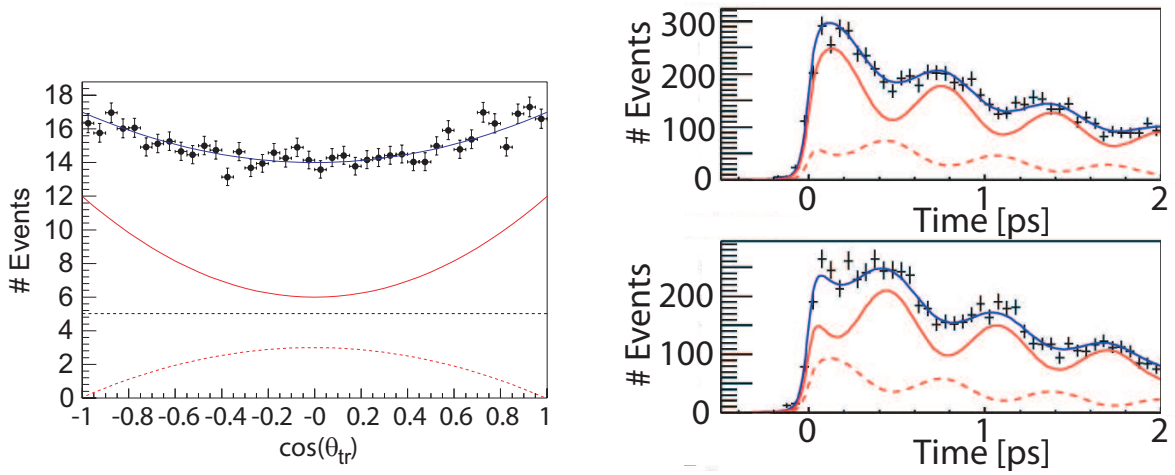


Figure 6.13: On the left hand side a projection of the likelihood fit on the $\cos\theta_{tr}$ *contribution is shown. On the right hand side the oscillations of* B_s *(top) and* $\overline{B_s}$ *(bottom) are shown for a value of* $\sin(2\delta\gamma)$ *which is 10 times larger than nominal. The curves through the data points are the projected sums of all contributions. The other solid curves are the CP even components, the dashed curves are the CP odd components and the dashed line in the plot on the left hand side corresponds to the background.*

The results of the 1000 fits can be converted to Gaussian distributions around the input values. These so-called pull distributions provide a measure for the accuracy of the 'toy experiments'. If the toy experiments are sufficiently realistic, these results also provide the sensitivity of the real experiment for the parameters of the likelihood fit. The results of

the studies presented in Ref. [82] are presented in Table 6.3. These results are based on 100,000 signal events with a purity of 87%. Based on the results of the selection presented in section 6.2.4, the number of signal events per year is estimated at $1.8 \cdot 10^6 \cdot 0.1 \cdot 0.87 \cdot 0.69 = 108,000$, with an efficiency of about 91%. It is expected that the errors on the results can be scaled to the new figures for efficiency and purity with

$$\frac{\left(\frac{signal}{\sqrt{signal+background}}\right)_{old}}{\left(\frac{signal}{\sqrt{signal+background}}\right)_{new}} \cdot \sigma_{old} \approx \frac{\left(\frac{100}{\sqrt{100+15}}\right)_{old}}{\left(\frac{108}{\sqrt{108+9}}\right)_{new}} \cdot \sigma_{old} = 0.93 \cdot \sigma_{old}. \tag{6.6}$$

Table 6.3: Expected statistical precision on the physics parameters extracted from the $B_s \rightarrow J/\psi\phi$ decay.

$R_T = 0.200 \pm 0.005$	$\sin(2\delta\gamma) = 0.04 \pm 0.06$	$\Delta\Gamma_s/\overline{\Gamma_s} = 0.10 \pm 0.02$	$\overline{\Gamma} = 1.54 \pm 0.02\,\text{ps}$

The precisions on the physics parameters as shown in Table 6.3 are expected to be enhanced with about 7% with the optimized set of selection criteria. The expected precisions clearly demonstrate the capabilities of *LHC*b, surpassing any existing *b*-physics experiment. Based on these results (and many others) one can expect that *LHC*b will provide major contributions to our understanding of the Standard Model and beyond.

Chapter 7

Summary

The Standard Model with three quark families incorporates CP violation in the weak interaction by means of a difference between the mass and the weak eigenstates, as implemented via the complex unitary three-by-three CKM matrix. Six of the nine unitarity conditions of this matrix can be drawn as triangles in the complex plane, of which two are relevant for the B-meson system: the lengths of the sides of the triangles can be extracted by measuring decay rates, and the relative angles can be extracted from CP-asymmetries. The analysis of B_s-meson decays enables us to over-constrain the unitary triangles. If the internal consistency of the triangles fails this will be an indication of new physics.

The interference in the decays of B_s and $\overline{B_s}$ to their common CP-eigen final state $(J/\psi \rightarrow \mu^+\mu^-)$ $(\phi \rightarrow K^+K^-)$ provides a direct measurement of the angle $\delta\gamma$, which relates the two triangles mentioned above. The final state consists of two spin-1 particles, which means an angular analysis is required to distinguish the CP-even and CP-odd components. Complications are introduced by the fast oscillations between the B_s and the $\overline{B_s}$, diluting the time integrated asymmetry and thus making a time dependent analysis necessary.

The *LHC*b experiment is designed to study CP violation and other rare phenomena in B-meson decay with high precision. It is being installed at the Large Hadron Collider (LHC), a proton accelerator with superconducting magnets providing collisions with a center of mass energy of 14 TeV at a rate of 40 MHz. In these interactions B mesons will be produced copiously, mostly in either the forward or backward cone centered along the beam axis due to the broad longitudinal energy distribution of the gluons in the protons. *LHC*b is therefore designed as a single arm spectrometer with an angular coverage from 10 mrad to 300 (250) mrad in the bending (non-bending) plane.

The *LHC*b design luminosity is $2 \cdot 10^{32} cm^{-2} s^{-1}$, to limit the number of pp interactions per event. As a result, the events are relatively clean and easy to analyse. About $10^{12} B\overline{B}$ pairs are expected to be produced per year, *i.e.* 1 in about 160 interactions is expected to produce such a pair. The trigger is a crucial component of *LHC*b, since the branching ratios of the interesting decays are in the order of $10^{-5} - 10^{-6}$. Fortunately, B mesons leave a clear signature in the detector: due to their high average boost and long life time they typically travel for several millimeters before decaying into other particles. The *LHC*b trigger is capable of identifying this signature, based on the impact parameter of the tracks. Other important aspects of this experiment include the vertex and momentum resolutions. The vertex resolution should be sufficient to resolve the rapid oscillations of the B_s meson

and the momentum resolution is required for accurate reconstruction of invariant masses, allowing the reduction of the vast background. Particle identification, *e.g.* separating kaons from pions, is another important criterion. This feature enables *LHC*b to distinguish the interesting decay channels and to minimize the systematic errors and dilution of the results.

The trigger of *LHC*b is split in multiple levels. The level-0 trigger rejects high-multiplicity events and selects events that contain decay products with high transverse momentum, as expected from the high B mass. This reduces the data rate (initially equal to the LHC beam crossing rate of 40 MHz) to on average 1 MHz. The efficiency of the level-0 trigger for muon channels is about 90% and for the remaining channels about 50%. The level-1 trigger is a software trigger that identifies events with a secondary vertex, based on the information of the vertex detector. The level-1 trigger reduces the event rate to 40 kHz with an efficiency varying between 45% and 80% for the various channels. A fast pattern recognition program tries to reconfirm the decision of the level-1 trigger in the higher level trigger (HLT). The most important decay channels are identified via exclusive selections, which might even include a partial RICH algorithm. The rest of the channels will be identified with inclusive selections. The output rate of the HLT is adjustable, but is expected to be 2 kHz at most. The efficiency of the HLT is on average about 95%, resulting in a combined trigger efficiency for the decay $B_s \rightarrow J/\psi\phi$ of about 70% (60%) in case of $J/\psi \rightarrow \mu^+\mu^-$ ($J/\psi \rightarrow e^+e^-$).

The tracking system consists of the vertex detector, the trigger tracker, a magnet and the outer and inner tracker system. All tracking stations are equipped with silicon sensors, except for the outer tracker, which uses straw tubes. The most interesting tracks for physics analysis are the long tracks, which traverse all tracking stations. In a B-event with this kind of tracks the decay length of a B-meson can be determined with an average resolution of about 22 μm in the transverse and 144 μm in the longitudinal direction. In combination with an average momentum resolution of 0.4%, this means that B_s-oscillations can be resolved with 5 σ sensitivity up to a mass difference of 68 ps^{-1} within 1 year of data taking.

Particle identification utilizes the ring imaging Cherenkov (RICH) detectors, calorimeters and muon chambers. The calorimeters and the muon chambers are included in the level-0 trigger decision for the identification of high momentum electrons, photons, pions, hadrons and muons. The particle identification algorithms in the offline analysis are based on likelihood estimators from all the particle identification detectors. The best separation of electrons, pions, muon, kaons and protons can be achieved by optimizing *e.g.* the sum of the log of the likelihood values for a specific decay. With this method 90% muon efficiency was obtained for the $B_s \rightarrow J/\psi\phi$ with less than 1% pion misidentification. Furthermore, the efficiency for kaons was 90% with about 5% pion misidentification.

The silicon sensors of the vertex detector are installed in a secondary vacuum only 8 mm from the beams. The small inner radius leads to short track extrapolation distances and thus allows for precise impact parameter reconstruction. During injection, the minimum required aperture is 30 mm, so the sensors will be retractable. A 250 μm thick aluminum foil shields the sensors from RF radiation from the beams. It also acts as a wakefield guide and avoids pollution of the beam vacuum. The thickness and shape of the foil are optimized to minimize the average amount of material that particles traverse. The foil can withstand a pressure difference of about 15 mbar, which implies that the pressures in the detector vacuum and in the beam vacuum need to be monitored and controlled at all times, especially during venting and evacuation.

Summary

The vertex detector of *LHC*b will use 300 μm thick n $-$ on $-$ n silicon micro-strip sensors. The signal in these sensors is mainly formed by the collection of electrons, which means that short pulses are produced. Moreover, the effects of radiation damage on the signal shape from these sensors are limited. The temperature of the sensors is maintained at -5°C by evaporation of CO_2 in a closed-loop cooling system, which reduces the radiation damage effects even further.

The sensors of the vertex detector as well as all other silicon detectors of *LHC*b are read-out by 128 channel analog chips called "Beetles". The chips are based on 0.25 μm CMOS integrated circuit technology and their radiation hardness is demonstrated up to 300 kGy. They run at 40 MHz and the output pulses are short enough to separate the information from consecutive bunch crossings. The maximum trigger latency is 4 μs, to match the requirements of the level-0 trigger. Furthermore, the chips are able to deliver full analog event information at 1 MHz, the input rate of the level-1 trigger.

An early version of the chip, the Beetle1.1, was extensively tested at the SPS accelerator facility at CERN. A prototype hybrid was equipped with 16 Beetle1.1 chips and a 300 μm thick p $-$ on $-$ n silicon sensor was used as detector. Furthermore, two scintillators were installed to generate a trigger signal and a separate XY station was used for tracking purposes. For the majority of the data, a single trigger issued the read-out of 8 consecutive time samples, spanning a total period of 200 ns. In this way the complete pulse shape was reconstructed from the data, without the need to change the latency. High statistics measurements demonstrated large variations in pulse shape characteristics within the chips as a function of channel number: the pulses from the low-numbered channels are faster and smaller than those from the high-numbered channels. This phenomenon has been baptized 'sticky charge' and is explained by trapped charge in a a capacitance of the read-out amplifier circuit. The problem has been solved in the Beetle1.3 and beyond.

The three most interesting parameters that characterize the pulse shape are rise time, signal/noise ratio and spill-over. The preferred method to determine the pulse shape is via reconstruction of cluster charges as a function of the time difference with the trigger. This method is cross-checked both with a method that uses the signal/noise ratio of a single-strip and with a method based on track selection. The comparison with the track selection analysis indicates a possible overestimate of about 0.7 in the signal/noise ratio for the cluster charge analysis, which is included in the systematic error.

The pulse shape characteristics of the Beetle1.1, tuned for *LHC*b, are:

- Signal/noise ratio = $17.5 \pm 0.9^{+0.1}_{-0.9}$
- Rise time = $23.5 \pm 0.7 \pm 0.5$ ns
- Spill-over = $36.1 \pm 1.3 \pm 1\%$.

Combined with the ENC relation of 871 e^- + 41.5 e^-/pF, this results in a predicted range of signal/noise ratios between 13.5 - 17.5 for the various sensor areas of the *LHC*b vertex detector. 'Single time-sample' data and 'high trigger rate' data are analyzed and found to be in good agreement with the rest of the results, except for a low signal/noise ratio for the high trigger rate analysis. This could be caused by interference from the test pulse circuit, but more measurements are required to draw definite conclusions on this issue. An analysis

of efficiency and spill-over versus threshold shows that the Beetle can operate at about 97% efficiency with 10% spill-over hits.

After the online selection the data flow of *LHC*b is estimated to be about 3 Petabyte per year. The computing power required to store and process these data, exceeds the resources available at any one of the participating research labs and universities. Large-scale distributed computing environments that cross administrative boundaries, also known as grids, are under construction to provide the solution to this problem. At the moment these grids already provide the means for the collaborations to produce millions of simulated events that mimic the LHC particle collisions in so called data-challenges. These large scale Monte Carlo productions are ideal test cases for the more complex situation of data analysis.

The *LHC*b Monte Carlo production system, DIRAC, has a client/server architecture based on a computing infrastructure distributed among the collaborating institutes. The resource acquisition is accomplished in two different ways. In the traditional approach, site managers are responsible for the installation of the client side software on their batch pool. In the grid approach, DIRAC was deployed on the European Data Grid (EDG) testbed, making use of the EDG tools for the acquisition of computing resources.

DIRAC is a tailored application, specifically designed to run on distributed resources. It could therefore be deployed in a grid environment without major modifications. The adoption of grid technologies by individual physicists can significantly be accelerated when they are able to run their own analysis on these resources as well. However, this requires non-expert users to deploy grid-unaware applications with a minimal amount of effort.

Here, a two fold implementation is tested to achieve these goals. Firstly, users interact locally with a well-known batch-system interface. The batch-system acquires remote resources in such a way that they appear and act as if they are controlled by the interface, effectively hiding the complexities associated with the various administrative domains. Secondly, the application runs in a virtual environment that provides the illusion of a local system, independent of the actual location of the resources it uses.

As a realistic test, the BaBar Monte Carlo production system - SP5 - was deployed on the EDG testbed. In theory this application has the right structure for distributed computing. The initial data can be distributed to a number of machines. Once initialized, any processor can produce an arbitrary number of events. In practice, the complexity of the application and in particular its dependence on a centralized 'Objectivity' database make it difficult to deploy. Nevertheless, it was quite feasible to deploy SP5 on the EDG testbed with the techniques described above: the cost of initialization equals the cost of computing only 70 events. Given that a typical instance of SP5 processes 10,000 events, the cost of remote execution can be amortized across a large run.

The simulated data produced with the *LHC*b data challenges are used for performance studies, in which they are processed in a similar way to normal data, which is a key feature of the *LHC*b software. The first part of these studies is the optimization of selection criteria for individual physics channels. This optimization is done with separate samples of signal and background events. The simulated sample of $B_s \rightarrow (J/\psi \rightarrow \mu^+\mu^-)(\phi \rightarrow K^+K^-)$ contains about 750,000 events, of which 30,000 are used to optimize the selection criteria. The most dangerous background for these events consists of mis-identified B-decays, of which 10 million events are simulated. The following criteria are investigated to distinguish between signal and background:

Summary

- Mass windows of B_s , J/ψ and ϕ.
- Transverse momenta of kaons and muons.
- Momentum of the ϕ.
- Normalized χ^2 of the decay vertices.
- Particle identification via the log-likelihood ratio for muons versus pions and kaons versus pions.
- Proper lifetime of the B-meson.

All these criteria contribute significantly to the separation of signal events from background events, except for the normalized χ^2 of the ϕ vertex fit. The optimal performance is defined as the best ratio of $Signal/\sqrt{Signal + Background}$, which is equivalent to finding the maximum significant observation of the signal events.

The optimal performance is identified by scanning the complete parameter space of the mentioned (correlated) selection criteria. A fast algorithm is developed, based on N sorted views on the data, enabling the analysis of about 100,000 events with 5 correlated selection criteria (20 bins each) to complete in about 45 minutes on a 2 GHz machine. Two problems still prevent the combined optimization of all correlated selection criteria.

- The selection of $B_s \rightarrow J/\psi\phi$ decays contains more than 5 criteria, which means that the performance of the optimization algorithm is still insufficient. One solution would be to parallelize this algorithm and distribute it over a large number of machines in a grid environment. Another solution, pursued here, is to group the criteria with the strongest correlations and iteratively optimize several of these groups.
- The number of simulated background events is insufficient to optimize all criteria together. Enough background events will be available when *LHC*b becomes operational, but for now the optimization of the individual groups is decoupled, neglecting the correlations between them. The values found for the selection criteria need to be relaxed slightly, to compensate for this (incorrect) assumption and to increase the efficiency. The resulting selection has an efficiency of about 10% and the sample has a purity between 85% and 97% with 90% confidence in the estimate.

The numbers found with the selections serve as input for the toy Monte Carlo studies. In these studies, multidimensional probability density functions (PDFs) are constructed which are expected to mimic the spectra that will come from an analysis of data acquired by *LHC*b. These PDFs combine information about the physics of the processes with experimental effects to allow for a realistic prediction of the *LHC*b performance. In the case of $B_s \rightarrow J/\psi\phi$, the main inputs for the PDFs are the oscillatory behavior between the B_s and the $\overline{B_s}$, the angular distribution of the $J/\psi\phi$ decay, the mass distribution of the B_s as obtained with the selection, the mis-tag rate and resolution effects. A combined likelihood fit is performed on these data and repeated for 1000 different data sets. Each of these sets is the equivalent of 1 year of data taking with the *LHC*b experiment.

Note that Δm_s is provided as an external parameter and is not extracted from this fit, since the sensitivity for this parameter is insufficient due to the expected small value for $\sin(2\delta\gamma)$. The results of all variables match within the error margin with the specified input values ($R_T = 0.200 \pm 0.005$, $\sin(2\delta\gamma) = 0.04 \pm 0.06$, $\overline{\Gamma} = 1.54 \pm 0.02\,\text{ps}$ and $\Delta\Gamma/\overline{\Gamma} = 0.10 \pm 0.02$), which means that the toy Monte Carlo studies are internally consistent. The errors provide an indication of the sensitivity of the real experiment for these variables, assuming that the description of the experimental effects is sufficiently accurate.

In the introduction of this thesis the issue of absence of antimatter in our universe was raised. The Standard Model does not provide sufficient CP violation to explain the observed baryon asymmetry and new physics is required to describe this phenomenon correctly [84]. Since the expected value of $\sin(2\delta\gamma)$ is small, a direct measurement of this parameter might well result in the first discovery of an inconsistency in the CKM sector of the Standard Model. In that case, the measurements of the *LHC*b experiment will provide a unique contribution to the solution of this cosmological mystery.

Appendix A
PLC control schema

In this appendix details are given of one of the PLC procedures to control the vacuum system of the vertex detector. The procedure describes the transition from the state where the primary vacuum contains 1 bar of ultra pure neon and the secondary vacuum 1 bar of nitrogen or air to the state where the vertex detector is in operational mode.

Figure 7.1 gives a schematic overview of the vacuum components of the vertex detector. These components can also be recognized in the first column of the PLC control schema shown on the next pages. The described PLC procedure is divided in three stages. During the first stage (evacuation from atmospheric pressure to 1 mbar), the pressure difference over the thin foil between primary and secondary vacuum needs to be monitored and controled to avoid plastic deformations to this foil. The second stage contains the evacuation by the turbo pumps. Since the pressure is already below 1 mbar, the pressure diffence over the foil is no longer an issue in this stage. At the end of this stage the secondary vacuum reaches its operational pressure of $10^{-5}\,mbar$. The last stage is entered when the pressure in the primary vacuum is below $10^{-7}\,mbar$. At that moment evacuation of the primary vacuum is transfered from turbo pumps to ion getter pumps, in order to reach an operational pressure of $10^{-8}\,mbar$.

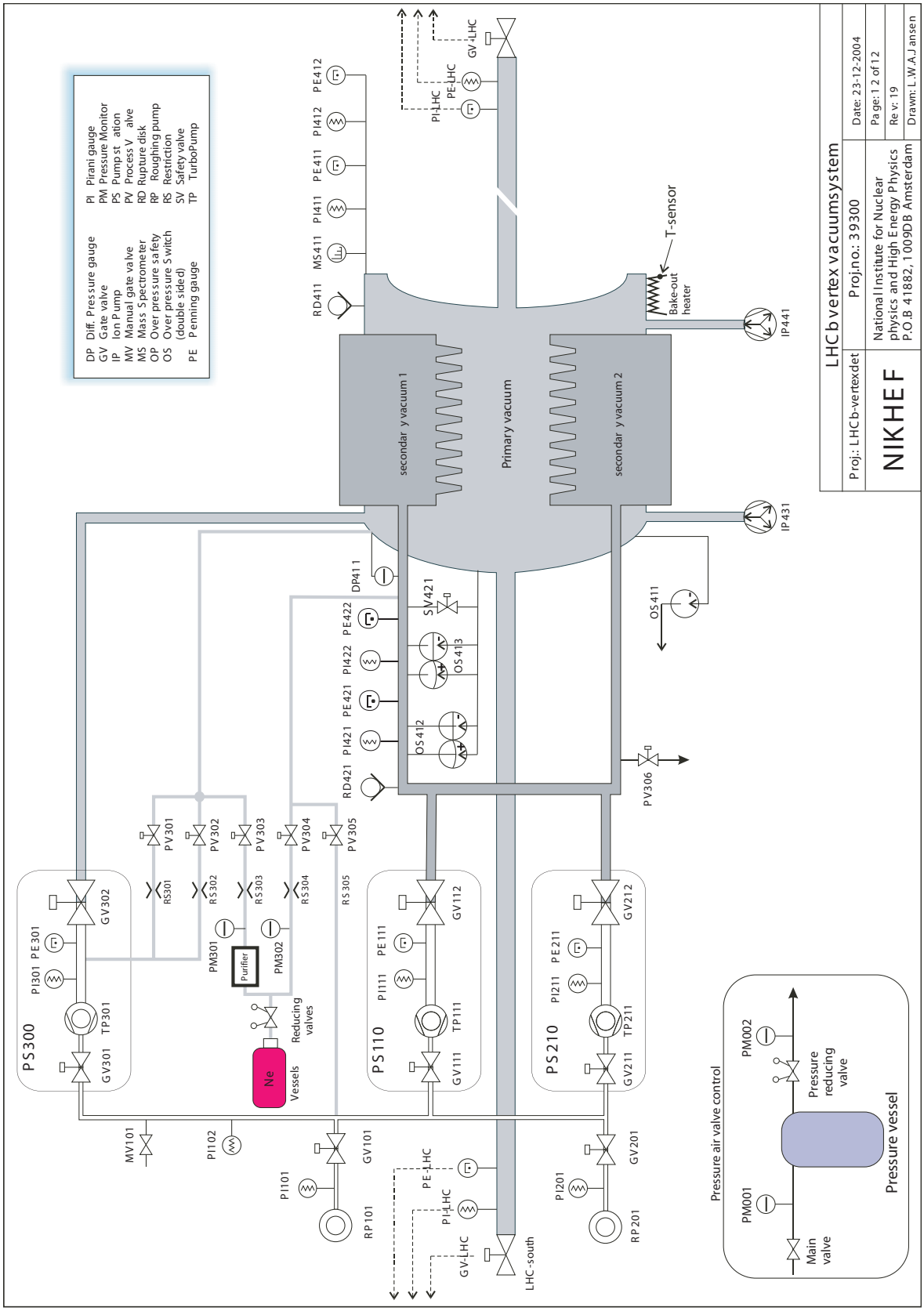


Figure 7.1: Schematic overview of the vacuum components of the vertex detector

	range	identifier	status	Evacuate from 1000 to 1 mbar	Evacuate from 1 to 10-7 mbar	Primary vacuum to 10-8 mbar	status
PI101	on-off	GaugeReg4	on				on
		AnP(0)	2,0E-1	**1,0E+3** ... **1,0E+0**	1,0E+0 ... **1,0E-3**		1,0E-3
PI102	on-off	GaugeReg3	on				on
		AnP(1)	2,0E-1	1,0E+3 ... 1,0E+0	1,0E+0 ... 1,0E-3		1,0E-3
PI201	on-off	GaugeReg2	on				on
		AnP(2)	2,0E-1	1,0E+3 ... 1,0E+0	1,0E+0 ... 1,0E-3		1,0E-3
PI111	on-off	GaugeReg1	on				on
		AnP(3)	2,0E-1		1,0E+0 ... 1,0E-3		1,0E-3
PE111	on-off	GaugeReg20	on	off	on		on
		AnP(4)	2,5E+0	**4,3E-3**	**1,6E-3** ... **1,0E-5**	... **1,3E-6**	1,0E-5
PI211	on-off	GaugeReg19	on				on
		AnP(5)	2,0E-1		1,0E+0 ... 1,0E-3		1,0E-3
PE211	on-off	GaugeReg18	on	off	on		on
		AnP(6)	2,5E+0	4,3E-3	1,6E-3 ... 1,0E-5	... 1,3E-6	1,0E-5
PI411	on-off	GaugeReg15	on				on
		AnP(9)	1,0E+1	1,0E+1 ... 1,0E+0	1,0E+0 ... 1,0E-3		1,0E-3
PE411	on-off	GaugeReg14	off		on		on
		AnP(10)	1,0E+1		1,6E-3 ... **1,2E-7**	... **1,3E-8**	1,2E-7
PI412	on-off	GaugeReg11	on				on
		AnP(13)	1,0E+1	1,0E+1 ... 1,0E+0	1,0E+0 ... 1,0E-3		1,0E-3
PE412	on-off	GaugeReg10	off		on		on
		AnP(14)	1,0E+1		1,6E-3 ... 1,2E-7	... 1,3E-8	1,2E-7
PI421	on-off	GaugeReg17	on				on
		AnP(7)	1,0E+1	1,0E+1 ... 1,0E+0	1,0E+0 ... **1,0E-3**		1,0E-3
PE421	on-off	GaugeReg16	off		on		on
		AnP(8)	1,0E+1		1,6E-3 ... 1,2E-7	... 1,3E-8	1,2E-7
PI422	on-off	GaugeReg13	on				on
		AnP(11)	1,0E+1	1,0E+1 ... 1,0E+0	1,0E+0 ... 1,0E-3		1,0E-3
PE422	on-off	GaugeReg12	off		on		on
		AnP(12)	1,0E+1		1,6E-3 ... 1,2E-7	... 1,3E-8	1,2E-7
PI301	on-off	GaugeReg9	on				on
		AnP(15)	2,0E-1	1,0E+3 ... 1,0E+0	1,0E+0 ... 1,0E-3	1,0E-3 ... 1,0E+3	1,0E+3
PE301	on-off	GaugeReg8	on	off	on	off	off
		AnP(16)	3,6E+0	4,3E-3	1,6E-3 ... 1,2E-7	1,2E-7 4,3E-3	4,3E-3
		ReedR(0)	0,0E+0				
OS411	5 mbar	ReedR(1)	closed				closed
OS412A	~~2 mbar~~	ReedR(2)	closed	open closed			closed
OS412B	2 mbar	ReedR(3)	open	closed open			open
OS413A	~~5 mbar~~	ReedR(4)	closed				closed
OS413B	5 mbar	ReedR(5)	open				open

PA301		AnP(20)	0							
PA302		AnP(21)	0							
DP411		AnP(22)	0							
RP101	on-off	RdCont(0)	0							
	status	ReedR(6)	on	**on**						on
RP201	on-off	RdCont(1)	0		RP101 or RP201 or both on					
	status	ReedR(7)	off	**off**	**inclusive OR**					0
TP111	standby on-off	RdCont(2)	0							
	motor on-off	RdCont(3)	0							
(0-10V	setspeed 0-10V	AnAdc(0)	0							
(0-10V)	statspeed	AnPump(0)	0							0%
	threshold	ReedR(8)	0					on		on
	errormess.	ReedR(9)	0							
TP211	standby on-off	RdCont(8)	0							
	motor on-off	RdCont(9)	0							
	setspeed 0-10V	AnAdc(2)	0							
	statspeed	AnPump(2)	0							0%
	threshold	ReedR(12)	0					on		on
	errormess.	ReedR(13)	0							
TP301	standby on-off	RdCont(11)	0							
	motor on-off	RdCont(12)	0							
	setspeed 0-10V	AnAdc(3)	0							
	statspeed	AnPump(3)	0							0%
	threshold	ReedR(14)	on					on	off	off
	errormess.	ReedR(15)	0							
			0							
IP431	on-off	RdCont(14)	0							
	status on-off	ReedR(16)	off						on	on
	fault 0-1	ReedR(17)	0							
	vacuum	AnP(23)	0						0 / 7	7
	Highvolt	AnP(24)	0						0 / 7	7
IP441	on-off	RdCont(16)	0							
	status on-off	ReedR(18)	off						on	on
	fault 0-1	ReedR(19)	0							
	vacuum	AnP(26)	0						0 / 7	7
	Highvolt	AnP(27)	0						0 / 7	7
GV101	open/close	AC24(0)	0							
	status	RdOp(0) RdCl(0)	open	**open**	closed when RP101=off					open
GV201	open/close	AC24(1)	0		**inclusive OR**					

	status	RdOp(1) RdCl(1)	closed	**closed**	closed when RP201=off			closed
GV111	open/close	AC24(4)	0					
	status	RdOp(4) RdCl(4)	open					open
GV112	open/close	AC24(5)	0					
	status	RdOp(5) RdCl(5)	closed			open		open
GV211	open/close	AC24(6)	0					
	status	RdOp(6) RdCl(6)	open					open
GV212	open/close	AC24(7)	0					
	status	RdOp(7) RdCl(7)	closed			open		open
GV301	open/close	AC24(8)	0					
	status	RdOp(8) RdCl(8)	open				closed	closed
GV302	open/close	AC24(9)	0					
	status	RdOp(9) RdCl(9)	closed			open	closed	closed
GV412	open/close	AC24(10)	0					
	status	RdOp(10) RdCl(10	closed					closed
SV421	open/close	AC24(11)	0					
	status	RdOp(11) RdCl(11	**closed**	closed				closed
PV301	open/close	AC24(12)	0					
	status	RdOp(12) RdCl(12	open				open	open
PV302	open/close	AC24(13)	0		closed open			
	status	RdOp(13) RdCl(13	closed	open		closed		closed
PV303	open/close	AC24(14)	0		closed open			
	status	RdOp(14) RdCl(14	closed	open		closed		closed
PV304	open/close	AC24(15)	0					
	status	RdOp(15) RdCl(15	closed					closed
PV305	open/close	AC24(16)	0					
	status	RdOp(16) RdCl(16	open					open
PV306	open/close	AC24(17)	0					
	status	RdOp(17) RdCl(17	closed	open				open
PV307	open/close	AC24(18)	0					
	status	RdOp(18) RdCl(18	open					open
PV308	open/close	AC24(19)	0					
	status	RdOp(19) RdCl(19	closed	open			closed	closed
PA001	open/close	AC24(3)	0					
	status	RdOp(3) RdCl(3)	open					open

= Simulated status = Simulated parameter open = PLC response on simulation abc = PLC action

PLC control schema

Bibliography

[1] S. Amatao *et al.*,"*LHC*b Technical Proposal", CERN/LHCC 98-4 (1998).

[2] R. Antunes Nobrega *et al.*, *"LHC*b Reoptimized detector", CERN/LHCC 2003-030 (2003).

[3] See *e.g.*, S. C. C. Ting *et al.*, "The Alpha Magnetic Spectrometer (AMS) on the International Space Stations, Results from the test flight on the Space Shuttle", Physics Reports **366** (2002) 331.

[4] A.D. Sakharov, "Violation of CP invariance, C asymmetry and baryon asymmetry of the universe", Sov. Phys. JETP Lett. **5** (1967) 24.

[5] A. Einstein, "Zur allgemeinen Relativitätstheorie", Sitzungsberichte der Königlich Preußische Akademie der Wissenschaften zu Berlin (1915) 778.

[6] L.H. Ryder, "Quantum Field Theory", Cambridge University Press, ISBN 0 521 478146, Second edition (1996).

[7] D. Gross and F. Wilczek, "Ultraviolet behaviour of non-abelian gauge theories", Phys. Rev. Lett. **30** (1973) 1343.

[8] S. Weinberg, "A model of leptons", Phys. Rev. Lett. **19** (1967) 1264.

[9] A. Salam, "Weak and Electromagnetic Interactions", Elementary Particle Theory: Relativistic Groups and Analyticity (1968) 367.

[10] F. Englert and R. Brout, "Broken symmetries and the mass of gauge vector mesons", Phys. Rev. Lett. **13** (1964) 321.

[11] P.W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett. **13** (1964) 508.

[12] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields", Nucl. Phys. **B** (1972) 189.

[13] S.L. Glashow, J. Iliopoulos, L. Maiani, "Weak interactions with lepton−hadron symmetry", Phys. Rev. **D** (1970) 1285.

[14] N. Cabibbo, "Unitary symmetry and leptonic decays", Phys. Rev. Lett. **10** (1963) 531.

[15] M. Kobayashi and T. Maskawa, "CP Violation in the renormalizable theory of weak interaction", Prog. Theor. Phys. **49** (1973) 652.

[16] J.H. Christenson *et al.*, "Evidence for the 2π decay of the K_2^0 meson, Phys. Rev. Lett. **13** (1964) 138.

[17] L. Wolfenstein, "Parametrization of the Kobayashi-Maskawa Matrix", Phys. Rev. Lett. **51** (1983) 1945.

[18] Y. Nir and H.R. Quinn, "Theory of CP violation in B Decays" in "B Decays", World Scientific (1994) 520.

[19] G. Branco, L. Lavoura and J. P. Silva, Oxford University Press, ISBN 0 19 850399 7 (1999).

[20] B. Aubert, *et al.*, "Measurement of the Branching Fractions and CP asymmetries of $B \to D^0 K^-$ decays with the BaBar detector", ICHEP (2004).

[21] K. Abe *et al.*, "Study of $B^\pm \to D K^\pm$ and $B^\pm \to D^* K^\pm$", ICHEP (2004).

[22] I. I. Bigi and A. I. Sanda, "CP violation", Cambridge University Press, ISBN 0 521 44349 0 (2000).

[23] A. J. Buras and Robert Fleischer, "Quark Mixing, CP violation and rare decays after the top quark discovery", hep-ph/9704376.

[24] N. Isgur and M. Wise, "Heavy quark symmetry", CEBAF-TH-92-10 (1992).

[25] P.F. Harrison and H.R. Quinn, "Physics at an Asymmetric B Factory", The BaBar Physics Book, SLAC Report 504 (1998).

[26] M. Ciuchini *et al.*, "Lifetime difference and CP violation parameters of neutral B mesons at the Next-to-leading order in QCD", hep-ph/0308029 (2003).

[27] J. Ellis, M.K. Gaillard and D.V. Nanopoulos, "A Phenomenological profile of the Higgs boson", Nucl. Phys. **B 106** (1976) 292.

[28] B. Aubert *et al.*, "Hadronic B decays with the BaBar detector", Phys. Rev. Lett. **87** (2001).

[29] M. E. Rose, "Elementary theory of angular momentum", John Wiley & Sons (1957).

[30] Suh Urk Chung, "Spin Formalisms", CERN 71-8 (1971).

[31] Isard Dunietz *et al.*, "How to extract CP-violating asymmetries from angular correlations", Phys. Rev. **D 43** (1991) 2193.

[32] K. Abe *et al.*, "Time-dependent Angular Analysis of B Decays", Belle note 419.

[33] A. S. Dighe *et al.*, "Angular distributions and lifetime differences in $B_s \to J/\psi\phi$ decays", FERMILAB-PUB-95/345-T (1995).

[34] Isard Dunietz, "$B_s - \overline{B_s}$mixing, CP violation and extraction of CKM phases from untagged B_s data samples, Phys. Rev. **D 52** (1995) 3048 .

[35] A. S. Dighe *et al.*, Extracting CKM phases and $B_s - \overline{B_s}$ mixing parameters from angular distributions of non-leptonic B decays, CERN-th/98-85 (1998).

[36] J. Charles *et al.*, CP violation and the CKM matrix: assessing the impact of the asymmetric B factories, hep-ph/0406184 v2 (2004).

[37] K. Hagiwara *et al.*, "The Review of Particle Physics", Physical Review **D66** (2002).

[38] A. Hocker *et al.*, "A new approach to a global fit of the CKM Matrix", Euro Physics **J.C 21**, (2001) 225.

[39] Y. Baconnier *et al.*, "The Large Hadron Collider Accelerator Project", CERN/AC/93-03 (1993).

[40] P. Lefèvre *et al.*, "LHC", CERN/AC/95-05 (1995).

[41] V. Baglin *et al.*, "Performance of a cryogenic vacuum system (COLDEX) with a LHC type proton beam", Vacuum **73** (2004) 201.

[42] S. Frixione *et al.*, "Heavy-quark production", Advanced Series on Directions in High Energy Physics **15** (1998) 609.

[43] E. Norrbin and T. Sjöstrand, "Production and Hadronization of Heavy Quarks", hep-ph/0005110 (2000).

[44] H. Dijkstra *et al.*, "Some remarks on systematic effects of the trigger and Event Generator studies", (internal) *LHC*b 2003-157 (2003).

[45] B. Andersson, "The Lund Model", ISBN 0-521-42094-6, Cambridge University Press (1998).

[46] N. Zaitsev, "Study of the *LHC*b pile-up trigger and the $B_s \to J/\psi\phi$ decay", CERN-THESIS-2000-043 (2000).

[47] R. Antunes Nobrega *et al.*, *"LHC*b trigger system", CERN/LHCC 2003-031 (2003).

[48] P. R. Barbosa Marinho *et al.*, *"LHC*b VELO", CERN/LHCC 2001-011 (2001).

[49] N. van Bakel *et al.*, "Wake fields in the *LHC*b Vertex detector", *LHC*b 99-043 and 99-044 (1999).

[50] J. van den Brand *et al.*, "The vacuum system of the *LHC*b vertex detector", Vacuum **67**, (2002) 363.

[51] C. J. S. Damerell, "Vertex detectors: The state of the art and future prospects", SSI (1995).

[52] L. Beattie *et al.*, "Charge collection efficiency in heavily irradiated silicon diodes", Nucl. Instr. and Meth. **A 412** (1998) 238.

[53] R. Brenner *et al.*, "RD-20 front-end architecture", Nucl. Instr. and Meth. **A 339** (1994) 564.

[54] D. Baumeister, "Development and Characterisation of a Radiation Hard Readout Chip for the *LHC*b-Experiment", ASIC laboratory of Heidelberg (2003).

[55] S. Klous *et al.*, "Characteristics of 16 Beetle1.1 chips on a VELO hybrid", *LHC*b 2003-069 (2003).

[56] Doris Eckstein, "The *LHC*b Vertex Locator", *LHC*b 2003-099 (2003).

[57] N. van Bakel *et al.*, "Investigation of the Beetle1.1 chip in the X7 testbeam", *LHC*b 2002-053 revised version (2002).

[58] M. Feuerstack-Raible *et al.*, Nucl. Instr. and Meth. **A 447** (2000) 35.

[59] P. Gray and R.G. Meyer, "Analysis and design of analog integrated circuits", ISBN 0-471-59984-0 (1993).

[60] M. Agari, "Studies of the Beetle 1.2 Pipeline Homogeneity", *LHC*b 2003-155.

[61] H. Bichsel, Rev. Mod. Phys. **60** (1988) 663.

[62] S. Hancock *et al.*, Phys. Rev. **A 28** (1983) 615.

[63] V. Blobel, "Formulae and methods in experimental data evaluation", Vol. 3 (1984).

[64] F.B. Hildebrand, "Introduction to numerical analysis", ISBN 0-486-65363-3 (1956).

[65] J. Libby, "A simulation study of clustering in the VELO testbeam software", LHCb 2002-036 (2002).

[66] "Cadence DFII Software (version IC 4.4.3) with 0.25 μm (cmos6sf) CMOS Library Design Kit version 1.0.2" (2001).

[67] E.N. Koffeman, Nucl. Instr. and Meth. **A 452** (2000) 89.

[68] H. Hoffman, "Report of the Steering Group of the LHC Computing Review", LHCC 2001-004.

[69] A. Tsaregorodtsev, *et al.*, "DIRAC - Distributed Infrastructure with Remote Agent Control", CHEP (2003).

[70] Heinz Stockinger *et al.*, File and Object Replication in Data Grids, Journal of Cluster Computing **5**, (2002) 305.

[71] J. Frey, T. Tannenbaum, I. Foster, M. Livny, and S. Tuecke, "Condor-G: A Computation Management Agent for Multi-Institutional Grids", Journal of Cluster Computing volume **5** (2002) 237.

[72] I. Foster and C. Kesselman, "Globus: A Toolkit-Based Grid Architecture", The Grid: Blueprint for a new Computing Infrastructure, Morgan Kaufmann, ISBN 1558604758 (1999) 259.

[73] Todd Tannenbaum, Derek Wright, Karen Miller, and Miron Livny, "Condor - A Distributed Job Scheduler", Beowulf Cluster Computing with Linux, The MIT Press, ISBN: 0-262-69274-0 (2002).

[74] S. Son and M. Livny, "Recovering Internet Symmetry in Distributed Computing", CCGrid (2003).

[75] S. Son and M. Livny, "Dealing with Internet Connectivity by an Extended Addressing Scheme", in preparation.

[76] D. Thain, S. Klous, and M. Livny, "Deploying Complex Applications in Unfriendly Systems with Parrot", Journal of Supercomputing (2004).

[77] I. Belyaev et al., "Simulation application for the *LHC*b experiment", CHEP (2003).

[78] I. Belyaev et al., "Integration of Geant4 with the Gaudi framework", CHEP (2001).

[79] Ellis, Stirling and Webber, "QCD and Collider physics", Cambridge University Press, ISBN 0-521-58189-3 (1996).

[80] B. Andersson, "The Lund Model", Cambridge University Press, ISBN 0-521-42094-6 (1998).

[81] Gerhard Raven, "Selection of $B_s \to J/\psi\phi$ and $B^+ \to J/\psi K^+$", *LHC*b 2003-118 (2003).

[82] Gerhard Raven, Sensitivity studies of χ and $\Delta\Gamma$ with $B_s \to (J/\psi \to \mu^+\mu^-)\ (\phi \to K^+K^-)$, LHCb 2003-119 (2003).

[83] W. Verkerke and D. Kirby, "The RooFit toolkit for data modeling", CHEP (2003).

[84] P. Huet and . Sather, "Electroweak baryogenesis and standard model CP violation", Phys. Rev. **D 51** (1995) 379.

Samenvatting

De schending van ladings- en pariteitsbehoud is ingebed in de electrozwakke wisselwerking van het Standaard Model met drie quark families door een verschil tussen de zwakke- en massaeigentoestanden. Dit verschil wordt gerepresenteerd in de complexe, unitaire drie bij drie CKM matrix. Zes van de negen unitariteitsvoorwaarden van deze matrix kunnen als driehoeken worden getekend in het complexe vlak. Twee van de zes driehoeken zijn relevant voor het B-meson systeem: de lengtes van de zijden van de driehoeken kunnen worden bepaald door het meten van vervalsamplitudes en de hoeken uit de CP-asymmetrieën. Door middel van de analyse van het verval van B_s-mesonen kunnen we de unitariteitsdriehoeken overbepalen. Als de driehoeken daardoor niet langer intern consistent zijn, is dat een indicatie voor verschijnselen die niet door het Standaard Model beschreven worden.

Een directe meting van de hoek $\delta\gamma$ wordt mogelijk met de interferentie tussen het verval van B_s en dat van $\overline{B_s}$ naar hun gemeenschappelijke CP-eigen eindtoestand $(J/\psi \to \mu^+\mu^-)$ $(\phi \to K^+K^-)$. Deze hoek relateert de twee eerder genoemde unitariteitsdriehoeken aan elkaar. De eindtoestand bestaat uit twee spin-1 deeltjes en dat maakt een hoekanalyse van dit verval noodzakelijk om de CP-even en CP-oneven componenten te kunnen scheiden. Verder is er een gecompliceerde tijdsafhankelijke analyse nodig omdat de snelle oscillaties tussen het B_s en het $\overline{B_s}$ de tijdsgeïntegreerde asymmetrie laten vervagen.

Het *LHC*b experiment is ontworpen om met hoge nauwkeurigheid CP schending en andere zeldzame fenomenen in het verval van B-mesonen te kunnen bestuderen. Het experiment wordt op dit moment geïnstalleerd in de 'Large Hadron Collider' (LHC), een proton (p) versneller op CERN in Genève met supergeleidende magneten, waar botsingen met een invariante massa van 14 TeV zullen plaatsvinden met een frequentie van 40 MHz. Er zullen vele B mesonen worden geproduceerd in deze interacties, meestal in de voorwaartse of achterwaartse richting langs de bundelas door de brede longitudinale energieverdeling van de gluonen in de protonen. *LHC*b is daarom ontworpen als een enkelarmige spectrometer met een hoekbereik van 10 tot 300 (250) mrad in het dispersieve (niet-dispersieve) vlak.

De ontwerpluminositeit van *LHC*b is $2 \cdot 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, zodat het aantal pp interacties per gebeurtenis beperkt zal blijven. Hierdoor zijn de gebeurtenissen relatief schoon en makkelijk te analyseren. Er zullen ongeveer $10^{12}\,B\overline{B}$ paren worden geproduceerd per jaar, met andere woorden men verwacht dat ongeveer 1 op de 160 pp interacties een dergelijk paar zal produceren. De trigger is een cruciale component van *LHC*b, aangezien de vervalsfracties van de interessante vervallen in de orde van 10^{-5} tot 10^{-6} zijn. Gelukkig laten B mesonen karakteristieke sporen achter in de detector: over het algemeen leggen ze een afstand af van een aantal millimeters door hun hoge boost en lange levensduur. Deze eigenschap wordt door de *LHC*b trigger herkend door middel van het bepalen van de kortste afstand tussen de deeltjessporen en de primaire vertex. Andere belangrijke aspecten van dit experiment zijn de

nauwkeurigheid in de vertex- en impulsbepaling. De vertex moet met voldoende oplossend vermogen worden bepaald in verband met de snelle oscillaties van de B_s mesonen en voldoende impulsresolutie is nodig voor nauwkeurige reconstructie van de invariante massa's, zodat de enorme achtergrond kan worden gereduceerd. Deeltjesidentificatie, in dit experiment in essentie het scheiden van kaonen en pionen, is een ander belangrijk criterium. Met behulp van dit kenmerk kan *LHC*b de interessante vervalskanalen onderscheiden en systematische fouten beperken.

De trigger van LHCb is verdeeld in verschillende niveaus. De niveau-0 trigger weigert gebeurtenissen met een hoog deeltjes aantal (multipliciteit) en selecteert gebeurtenissen waarbij vervalsprodukten een hoge impuls loodrecht op de bundelrichting hebben, zoals wordt verwacht door de hoge B massa. Hiermee wordt de gegevensstroom (die oorspronkelijk gelijk was aan de bundelfrequentie van 40 MHz) gereduceerd tot gemiddeld 1 MHz. De niveau-0 trigger is ongeveer 90% efficiënt voor de kanalen met muonen in de eindtoestand en ongeveer 50% voor de overige kanalen. De niveau-1 trigger is een software trigger die secundaire vertices identificeert met gebruik van de informatie van de vertex detector. Deze trigger reduceert de gegevensstroom naar 40 kHz, met een efficiëntie die varieert van 45% tot 80% voor de verschillende vervalskanalen. Een snel patroonherkenningsalgoritme probeert de beslissing van de niveau-1 trigger te bevestigen in de hogere niveau trigger. De belangrijkste vervalskanalen worden vervolgens herkend via exclusieve selecties, die zelfs een gedeeltelijk 'ring imaging Cherenkov' (RICH) algoritme kunnen bevatten. De rest van de vervalskanalen wordt geïdentificeerd met inclusieve selecties. De uitgangsfrequentie van de hogere niveau trigger zal naar verwachting niet hoger zijn dan 2 kHz. De hogere niveau trigger is gemiddeld ongeveer 95% efficiënt, hetgeen resulteert in een gecombineerde trigger efficiëntie voor $B_s \to J/\psi\phi$ van ongeveer 70% (60%) in het geval van $J/\psi \to \mu^+\mu^-$ ($J/\psi \to e^+e^-$).

Het spoorreconstructiesysteem bestaat uit de vertex detector, de 'trigger tracker', een magneet en het 'outer tracker' en 'inner tracker' systeem. Alle spoorreconstructie detectoren zijn uitgerust met silicium sensoren, behalve de 'outer tracker', die gebruik maakt van rietjes. De interessantste sporen voor de fysica zijn de zogenaamde 'lange sporen' van deeltjes die door alle spoorreconstructiedetectoren vliegen. In gebeurtenissen met dergelijke 'lange sporen' is de gemiddelde resolutie voor de vervalslengte van een B deeltje 22 μm in de loodrechte en 144 μm in de longitudinale richting. Dit betekent dat de limiet van 5 σ gevoeligheid voor B_s-oscillaties wordt bereikt bij een massa verschil tussen het B_s en $\overline{B_s}$ meson (Δm_s) van 68 ps^{-1} als er 1 jaar wordt gemeten met het *LHC*b experiment.

De deeltjesidentificatie maakt gebruik van de RICH detectoren, calorimeters en muon kamers. De calorimeters en de muon kamers worden ook gebruikt voor de niveau-0 trigger beslissing om elektronen, fotonen, pionen, hadronen en muonen te identificeren met een hoge impuls loodrecht op de bundelrichting. De deeltjesidentificatiealgoritmes in de offline analyse zijn gebaseerd op likelihood schattingen van alle deeltjesidentificatie detectoren. De beste scheiding van elektronen, pionen, muonen, kaonen en protonen kan worden bereikt door bijvoorbeeld de som van de logaritme van de likelihood waarden te optimaliseren voor een specifiek verval. Deze methode leidde tot een selectie voor $B_s \to J/\psi\phi$ met een efficiëntie van 90% voor muonen, terwijl minder dan 1% van de selectie bestond uit foutief geïdentificeerde pionen. Verder was de efficiëntie voor kaonen 90% met ongeveer 5% van de selectie bestaande uit foutief geïdentificeerde pionen.

De silicium sensoren van de vertex detector zijn geïnstalleerd in een secundair vacuüm op slechts 8 mm van de bundels. Deze kleine afstand leidt tot korte extrapolatieafstanden voor de sporen en daarom kan een nauwkeurige reconstructie worden gemaakt van de kortste afstand tussen de primaire vertex en een spoor. Tijdens het injecteren van de bundels moet de minimale opening 30 mm zijn, dus de sensoren moeten kunnen worden teruggetrokken. Een 250 μm dik aluminium folie beschermt de sensoren tegen radiofrequentie straling van de bundels. Dit folie werkt ook als wakefield geleiding en voorkomt vervuiling van het bundelvacuüm. De dikte en vorm van het folie zijn zo geoptimaliseerd dat deeltjes door een minimale hoeveelheid materiaal vliegen. Het folie kan een drukverschil van 15 mbar weerstaan. Dat betekent dat de drukken in zowel het detector- als het bundelvacuüm continu moeten worden gemeten en geregeld, in het bijzonder tijdens het beluchten en het afpompen.

De vertex detector van *LHC*b zal 300 μm dikke n $-$ on $-$ n silicium micro-strip sensoren gebruiken. De signalen van deze sensoren worden voornamelijk gevormd door elektronen, waardoor korte pulsen kunnen worden geproduceerd. Bovendien zijn de effecten van stralingsschade op de signaalvorm van deze sensoren beperkt. De sensoren worden gekoeld op een temperatuur van -5°C door verdamping van CO_2 in een gesloten circulair systeem. Hierdoor wordt stralingsschade nog verder beperkt.

Zowel de sensoren van de vertex detector als de andere silicium detectoren in *LHC*b worden uitgelezen door 128 kanaals analoge chips die 'Beetles' heten. Deze chips zijn gebaseerd op 0.25 μm CMOS 'integrated circuit' technologie en hun stralingshardheid is aangetoond tot een geïntegreerde dosis van 300 kGy. Ze werken op 40 MHz en de uitgangspulsen zijn kort genoeg om de informatie van opeenvolgende bundelinteracties te kunnen onderscheiden. De maximale beschikbare triggervertraging van deze chip is 4 μs, in overeenstemming met de specificaties van de niveau-0 trigger. Verder is de chip in staat om de volledige analoge informatie van de gebeurtenissen aan te leveren op een frequentie van 1 MHz, gelijk aan de ingangsfrequentie van de niveau-1 trigger.

Een van de eerste versies van de chip, de Beetle1.1, is uitgebreid getest in de SPS versneller faciliteit op CERN. Een prototype hybride is uitgerust met 16 Beetle1.1 chips en een 300 μm dikke p $-$ on $-$ n silicium sensor is als detector gebruikt. Verder waren er twee scintillatoren geïnstalleerd om de trigger te kunnen genereren en werd een apart XY station gebruikt voor spoorreconstructie. Voor de meerderheid van de gemeten gegevens gold dat een trigger leidde tot het uitlezen van 8 opeenvolgende ingangssignalen, zodat een totale periode van 200 ns werd overbrugd. Op deze manier kon de volledige pulsvorm worden gereconstrueerd zonder dat het nodig was om de triggervertraging te veranderen. Metingen met hoge statistiek lieten grote variaties zien in de pulsvorm karakteristieken binnen de chips, als functie van het kanaal nummer. Pulsen van kanalen met lage nummers zijn sneller en kleiner dan die van kanalen met hoge nummers. Dit fenomeen is 'sticky charge' gedoopt en wordt veroorzaakt door gevangen lading in een capaciteit van het uitleesversterkercircuit. Het probleem is opgelost in de Beetle1.3 en latere versies van de chip.

De drie interessantste parameters die de pulsvorm karakteriseren zijn stijgtijd, signaal/ruis verhouding en 'spill-over'. De pulsvorm is bij voorkeur bepaald met een methode die gebruik maakt van ladingsclusters als functie van het tijdsverschil met de trigger. Deze methode is vervolgens gecontroleerd met zowel een methode die de signaal/ruis verhouding van een enkele strip gebruikt, als met een methode gebaseerd op spoorselectie. Een mogelijke overschatting van de signaal/ruis verhouding met ongeveer 0.7 kwam aan het licht door de

vergelijking met de spoorselectie analyse. Deze bijdrage is verwerkt in de systematische fout.

De pulsvorm karakteristieken voor de Beetle1.1, geoptimaliseerd voor *LHC*b, zijn:

- Signaal/ruis verhouding = $17.5 \pm 0.9^{+0.1}_{-0.9}$
- Stijgtijd = $23.5 \pm 0.7 \pm 0.5$ ns
- Spill-over = $36.1 \pm 1.3 \pm 1\%$.

Door deze getallen te combineren met de relatie voor de 'equivalent noise charge' (ENC) kan de voorspelling worden gedaan dat met deze hardware de signaal/ruis verhoudingen tussen de 13.5 - 17.5 zullen zijn voor de verschillende sensor gebieden in de *LHC*b vertex detector. 'Enkel ingangssignaal' gegevens en 'hoge-trigger-frequentie' gegevens zijn ook geanalyseerd en ze waren goed in overeenstemming met de overige resultaten. Alleen de signaal/ruis verhouding van de 'hoge-trigger-frequentie' analyse was erg laag. Dit zou kunnen worden veroorzaakt door interferentie vanuit het testpuls circuit, maar er zijn meer metingen nodig om hier definitieve conclusies over te trekken. Door de efficiëntie en de spill-over uit te zetten als functie van de drempelwaarde, is aangetoond dat de Beetle kan worden bedreven met 97% efficiëntie, als wordt geaccepteerd dat 10% van de spill-over de drempel passeert.

Na de selectie door de triggers zal de gegevensstroom die uit het *LHC*b experiment komt ongeveer 3 Petabyte per jaar zijn. Het aantal computers dat nodig is om deze gegevensstroom te verwerken en op te slaan is veel groter dan ieder van de deelnemende onderzoekslaboratoria en universiteiten afzonderlijk kan leveren. Gedistribueerde computerinfrastructuren, die over administratieve domeinen heen reiken, zijn op dit moment in ontwikkeling om dit probleem te kunnen oplossen. Deze infrastructuren zijn ook wel bekend als 'grids'. Op dit moment leveren deze 'grids' al de middelen voor de samenwerkingsverbanden om miljoenen gebeurtenissen te simuleren in zogenaamde 'gegevens-challenges'. Deze gebeurtenissen moeten de karakteristieken van de LHC interacties nabootsen. Deze grootschalige Monte Carlo producties vormen de ideale tests voor de nog gecompliceerdere situatie van de gegevens analyse.

Het *LHC*b Monte Carlo productie systeem, DIRAC, heeft een 'client/server' architectuur, gebaseerd op een over de deelnemende instituten gedistribueerde computerinfrastructuur. In dit systeem kan het verwerven van computers op twee verschillende manieren worden bereikt. Op de traditionele manier zijn managers op ieder instituut verantwoordelijk voor het installeren van de 'client software' op hun 'batch-systeem'. In het geval van een grid, DIRAC is geïnstalleerd en gebruikt op het 'European gegevens Grid (EDG) testbed', worden de EDG hulpmiddelen gebruikt om de benodigde computers te verwerven.

DIRAC is speciaal ontwikkeld om op gedistribueerde systemen te kunnen draaien. Het kon daarom zonder veel aanpassingen worden gebruikt in een grid omgeving. Het overnemen van grid technologieën door individuele fysici zou behoorlijk kunnen worden versneld als zij in staat zouden zijn om deze infrastructuur voor hun eigen analyses te gebruiken. Echter, dit vereist dat niet-experts hun programma's, die niet zijn ontwikkeld voor een gedistribueerde omgeving, met minimale inspanning kunnen gebruiken.

In dit proefschrift wordt een tweeledige implementatie getest die deze doelen moet kunnen realiseren. Ten eerste communiceren gebruikers met een lokale interface naar een 'batch-systeem'. Dit 'batch-systeem' verwerft computers op andere locaties, maar presenteert ze

alsof ze volledig worden gecontroleerd door dezelfde interface. Hierdoor worden complicaties gerelateerd aan het overschrijden van administratieve domeinen verborgen voor de gebruiker. Ten tweede draaien de programma's van de gebruiker in een virtuele omgeving, die ze voorziet van de illusie van een lokaal systeem, onafhankelijk van de werkelijke locatie van de computersystemen die worden gebruikt.

Als realistische test is het BaBar Monte Carlo productie systeem - SP5 - geïnstalleerd en gebruikt op het 'EDG testbed'. In theorie heeft dit programma de goede structuur voor gedistribueerd gebruik. De gegevens, nodig om op te starten, kunnen worden gedistribueerd over een aantal machines. Als deze machines zijn geïnitialiseerd, kan iedere processor een willekeurig aantal gebeurtenissen simuleren. In de praktijk maakt de complexiteit het moeilijk om deze applicatie te distribueren, in het bijzonder door de afhankelijkheid van een centrale 'Objectivity gegevensbase'. Toch was het goed mogelijk om SP5 op het 'EDG testbed' te installeren en te gebruiken met de eerder beschreven technieken: de initialisatie duurde ongeveer even lang als het berekenen van 70 gebeurtenissen. Typisch worden er per proces ongeveer 10,000 van deze gebeurtenissen gesimuleerd, dus de initialisatie levert slechts een minimale bijdrage aan de totale verwerkingstijd.

De in de 'gegevens challenges' geproduceerde gebeurtenissen worden gebruikt om de verwachte prestaties van *LHC*b te bestuderen. Een belangrijk aspect van de LHCb software is daarbij dat deze gesimuleerde gebeurtenissen op vergelijkbare manier worden behandeld als de echte gebeurtenissen. Het eerste deel van deze studies is het optimaliseren van de selectiecriteria voor individuele vervalskanalen. Deze optimalisatie is uitgevoerd met gescheiden verzamelingen van signaal en achtergrond gebeurtenissen. Er waren 750,000 gesimuleerde signaal gebeurtenissen beschikbaar voor het kanaal $B_s \rightarrow (J/\psi \rightarrow \mu^+\mu^-)(\phi \rightarrow K^+K^-)$, hiervan zijn er 30,000 gebruikt voor het optimaliseren van de selectiecriteria. De gevaarlijkste achtergrond voor dit signaal bestaat uit foutief geïdentificeerde B gebeurtenissen, waarvoor 10 miljoen kandidaten zijn geproduceerd. De volgende criteria zijn onderzocht om signaal en achtergrond van elkaar te kunnen onderscheiden:

- Massa snedes op B_s , J/ψ en ϕ.
- Snedes op de impuls loodrecht op de bundelrichting van kaonen en muonen.
- Een snede op de impuls van het ϕ.
- Snedes op de genormaliseerde χ^2 voor de fits van de vervalsvertices.
- Deeltjesidentificatie via de log-likelihoodverhouding voor muonen en pionen en voor kaonen en pionen.
- Levensduur in het rustsysteem van het B-meson.

Al deze criteria dragen significant bij aan de scheiding van signaal en achtergrond gebeurtenissen, behalve de genormaliseerde χ^2 van de ϕ vertex fit. De optimale prestatie is gedefinieerd als de beste verhouding $Signaal/\sqrt{Signaal + Achtergrond}$, wat gelijk staat aan het vinden van de maximaal significante observatie van signaal gebeurtenissen.

De optimale prestatie wordt gevonden door het scannen van de volledige parameterruimte van de genoemde (gecorreleerde) selectiecriteria. Een snel algoritme is ontwikkeld, gebaseerd op N gesorteerde projecties van de gegevens, die de analyse van 100,000 gebeurtenissen met 5 gecorreleerde selectiecriteria (elk 20 bins) uitvoert in ongeveer 45 minuten op een 2 GHz machine. Toch zijn er nog twee problemen die een gecombineerde optimalisatie van alle gecorreleerde selectiecriteria in de weg staan.

- De selectie van $B_s \to J/\psi\phi$ vervalsproducten bevat meer dan 5 criteria. Dat betekent dat de snelheid van het optimalisatiealgoritme nog te laag is. Het parallelliseren van dit algoritme, om het te kunnen verspreiden over een groot aantal machines in een grid omgeving, zou een oplossing zijn. Een andere oplossing, die hier gekozen is, is het groeperen van de criteria met de sterkste onderlinge correlaties. De groepen kunnen vervolgens iteratief worden geoptimaliseerd.
- Het aantal gesimuleerde achtergrondgebeurtenissen is onvoldoende om alle criteria tegelijk te kunnen optimaliseren. Er zal genoeg achtergrond beschikbaar zijn als *LHC*b operationeel wordt, maar op dit moment is de optimalisatie van de verschillende groepen ontkoppeld, waarbij de correlaties tussen de groepen worden verwaarloosd. De waarden die voor de selectiecriteria worden gevonden moeten vervolgens enigszins losser worden gekozen om te compenseren voor het (incorrect) verwaarlozen van de correlaties en om de efficiëntie te verhogen. De uiteindelijke selectie heeft een efficiëntie van 10% en de geselecteerde gebeurtenissen zijn tussen de 85% en 97% puur, waarbij de betrouwbaarheid van de schatting 90% is.

De waarden die bij de selecties zijn gevonden dienen als uitgangspunt voor de zogenaamde 'toy Monte Carlo' studies. In deze studies worden multidimensionale waarschijnlijkheidsdichtheidfuncties gemaakt, waarvan wordt verwacht dat ze de spectra nabootsen die uit een analyse van echte *LHC*b gegevens zullen komen. In deze waarschijnlijkheidsdichtheidfuncties wordt het fysisch model gecombineerd met experimentele effecten om de voorspellingen van de *LHC*b prestaties zo realistisch mogelijk te maken. In het geval van het verval $B_s \to J/\psi\phi$ bestaat de aanvangsinformatie uit het oscillerende gedrag van de B_s en de $\overline{B_s}$, de hoekverdeling van het $J/\psi\phi$ verval, de gereconstrueerde massa verdeling van de B_s zoals gevonden met de selecties, resolutie effecten, het percentage B_s mesonen dat als $\overline{B_s}$ wordt geïdentificeerd en andersom. Op de gegevens is een gecombineerde likelihood fit uitgevoerd en dat is herhaald voor 1000 verschillende sets. Iedere set is gelijkwaardig met de hoeveelheid gegevens die *LHC*b naar verwachting in 1 jaar meten beschikbaar heeft.

De waarde voor Δm_s wordt als externe parameter aangeboden en niet uit de fit gehaald. De gevoeligheid voor deze parameter is onvoldoende door de kleine verwachtingswaarde voor $\sin(2\delta\gamma)$. De resultaten van alle variabelen komen binnen de foutenmarges overeen met de gespecificeerde uitgangswaarden ($R_T = 0.200 \pm 0.005$, $\sin(2\delta\gamma) = 0.04 \pm 0.06$, $\overline{\Gamma} = 1.54 \pm 0.02\,\text{ps}$ en $\Delta\Gamma/\overline{\Gamma} = 0.10 \pm 0.02$). Dit betekent dat de 'toy Monte Carlo' studies intern consistent zijn. Als wordt aangenomen dat de experimentele effecten voldoende nauwkeurig zijn beschreven, dan geven de gespecificeerde fouten een indicatie van de gevoeligheid van het experiment voor deze variabelen.

In de inleiding van dit proefschrift werd het ontbreken van antimaterie in ons universum al even aangestipt. Het Standaard Model levert niet voldoende ladings- en pariteitsschending om de waargenomen baryonasymmetrie te kunnen verklaren. Er zijn dus verschijnselen nodig die niet in dit model voorkomen om dit fenomeen correct te kunnen beschrijven [84]. Aangezien de verwachtingswaarde van $\sin(2\delta\gamma)$ klein is, zou een directe meting van deze parameter de eerste ontdekking van een inconsistentie in de CKM sector van het Standaard Model kunnen opleveren. In dat geval leveren de metingen van het *LHC*b experiment een unieke bijdrage aan de oplossing van dit kosmologische mysterie.